Disagreement prompts young children's metacognitive reflection

By

Antonia Frederike Langenhoff

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jan Engelmann, Co-Chair
Professor Mahesh Srinivasan, Co-Chair
Professor Celeste Kidd
Professor Audun Dahl

Summer 2024

Disagreement prompts young children's metacognitive reflection

Abstract

Disagreement prompts young children's metacognitive reflection
By
Antonia Frederike Langenhoff
Doctor of Philosophy in Psychology
University of California, Berkeley
Professors Jan Engelmann & Mahesh Srinivasan, Co-Chairs

From Galileo to Gandhi, and from Plato to Piaget, influential thinkers throughout history have highlighted the benefits of disagreement for science, society, and the individual. Despite the rich theoretical interest, the specific individual benefits of disagreement have often remained unclear. To address this gap, the current dissertation explores one particular individual psychological consequence of disagreement: how it prompts metacognitive reflection during early childhood. The ability to metacognitively reflect on one's own knowledge plays a critical role in learning, as well as in individual and joint decision-making. Yet, young children's metacognitive capacities are often still limited in significant ways. By middle childhood, however, children's metacognitive competence has significantly improved. What explains this striking change? This dissertation argues that metacognitive development is centrally driven by young children's social experiences of disagreement.

To begin to test this hypothesis, the effects of disagreement on young children's metacognition are explored across three chapters, each focusing on a distinct, frequently studied dimension of metacognition: reason-giving, confidence ratings, and rational belief revision. Chapter 2 uses a cross-cultural approach, finding that experiencing disagreement (more so than agreement) leads children from three diverse cultural backgrounds to reflect on their reasons for their beliefs when making joint decisions. Chapter 3 demonstrates that disagreeing (versus agreeing) with another individual reduces young children's overconfidence and increases their motivation to search for the correct answer. Finally, Chapter 4 finds that children flexibly revise their initial beliefs, or suspend judgment until they have acquired additional evidence, depending on the strength of the evidence supporting their own belief versus that of a disagreeing other.

Together, these findings provide clear evidence that experiencing disagreement prompts young children's metacognitive reflection. Theoretically, these insights are significant as they bridge the gap between prior theoretical perspectives that have emphasized the role of social interaction in cognitive development, such as constructivist learning theories and cultural evolutionary accounts of metacognition, refines them, and makes them empirically testable. Moreover, the current work has important practical implications, and could inform interventions aimed at fostering learning and reasoning, as well as promoting mutual understanding.

Acknowledgements

With this dissertation, I wrap up my time as a graduate student at Berkeley, which is bittersweet. On the one hand, I'm sad to leave this creative and stimulating environment, and especially my two fantastic labs, the Social Origins Lab and the Language and Cognitive Development Lab. It is only with a little bit of irony when I say that if there was a sustainable way to be a grad student forever, I might just take it. On the other hand, I'm proud to have completed this degree, happy with how my work on disagreement has come together in this dissertation, and very much looking forward to my time as a Postdoc. One thing is clear: I couldn't have made it here without a lot of help. So, I want to give a big thanks to everyone who's been a part of this journey.

First, I want to thank my advisors, Jan Engelmann and Mahesh Srinivasan. Mahesh, thank you for taking a chance on me, from that initial internship to bringing me on as a grad student. I strive every day to be as patient, dedicated, and open to new ideas as you are. Jan, thanks for being such a great role model – I truly appreciate your passion for science, your willingness to question and revise your beliefs, and the sense of humor with which you approach your work. I am so lucky to have been (and continue to be) mentored by a team of advisors who are simultaneously as intelligent and caring as you both!

I am also deeply grateful to Audun Dahl. After we wrote one of my first-ever papers together, I feel like I managed to sneak my way into making you my key external mentor, and I'm incredibly thankful for your openness and willingness to step into this role. You always seemed to know exactly what to say, both to challenge me as an academic and to encourage me as a human. I also want to thank Celeste Kidd for being on my dissertation committee and for providing insightful feedback during my Qualifying exams.

To my dear lab members in the Social Origins and LCD Lab – as someone who studies collaborative reasoning, our discussions have always been prime examples of just that, and these exchanges have greatly contributed to my growth as a researcher. A special thank you to Oded Ritov, Josh Confer, and Colin Jacobs for sharing all the ups and downs of grad school with me, and for being excellent friends. Thank you to Rebecca Zhu for always being a step ahead and having the answer to any question I might have (and for coming by my desk and making me laugh). A big thanks also to Hanna Schleihauf, the best "academic big sister" I could have asked for. Your openness to collaborate, and your enthusiasm for and willingness to explain statistics have been so important to my time as a grad student (and I will never forget our wonderful research trip to Kenya).

The many incredibly talented and dedicated research assistants and summer interns who contributed their time and energy to this research deserve my sincere gratitude. Working with all of you has been one of the most rewarding aspects of graduate school. Special thanks also to lab managers Emily Chau, Alissa Gomez, and Eliza Swindell – this work could not have happened without you. Additionally, I want to extend my appreciation to the local preschools, schools, museums, and the Oakland Zoo, as well as their teachers, staff, and parents for their invaluable support and continued collaboration.

To my parents, thank you for instilling in me a profound love for curiosity, critical thinking, and learning. Your guidance, encouragement, and trust have made me the person I am today. And to my sister, who graciously accepted that I move far away for grad school (under the condition that I always come home for Christmas)—thank you. You are the best!

To my Californian and German friends: Thank you for the great conversations, for the laughter, and for the adventures. You've all made me feel at home in two places at once. I feel so grateful to have so many wonderful people in my life.

And finally, a huge thanks to Kurt. When I unexpectedly ended up in Berkeley's exchange semester program in 2016, I never imagined it would lead to spending the rest of my life with you. Thank you for your steadfast support of my research, my dreams, and me as a person. I'm truly fortunate to have you as my best friend and partner.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1 The benefits of disagreement

"Conflict is the gadfly of thought. It stirs us to observation and memory. It instigates to invention. It shocks us out of sheep-like passivity, and sets us at noting and contriving…conflict is a sine qua non of reflection and ingenuity."
John Dewey (1922)

While irreconcilable disagreements may cause polarization and social fragmentation, productive disagreements can propel science and society forward (see, e.g., Arendt, 1958; Habermas, 1962, 1997; Kuhn, 1962; Mill, 1859; Moshman, 2020; Sen, 2003; Westrum, 1989). Take Galileo's famous disagreement with the proponents of the then prevailing Aristotelian theory of gravity, for instance. By challenging the assumption that heavier objects fall faster than lighter ones, it laid the foundation for modern physics and astronomy (Lamers et al., 2021). Or consider the public debate about climate change, which has played a crucial role in shaping climate policies at local, national, and international levels (Dessler & Parson, 2019).

In addition to these collective-level benefits, disagreement – i.e., encounters with conversation partners holding conflicting views – also offers significant advantages at the individual psychological level. Specifically, constructivist learning theorists have long argued that the experience of disagreement can promote cognitive development (e.g., Dewey, 1922; Mercer & Littleton, 2007; Nussbaum, 2008; Piaget, 1952; Perret-Clermont, 1980; Vygotsky, 1987). Piaget, for example, recognized that "it is precisely by a constant interchange of thought with others that we are able to decentralize ourselves... to coordinate internal relations deriving from different view points" (1950, p.64).

Empirically, this idea is supported by a substantial body of research showing beneficial effects of disagreement on children's learning and reasoning abilities (e.g., Ames & Murray, 1982; Bearison et al., 1986; Doise et al., 1976; Doise & Mugny, 1978; Doise & Mugny, 1979; Felton et al., 2009; Fawcett & Garton, 2005; Schwarz et al., 2000; Sampson & Clark, 2009; Zohar & Nemet, 2002; for a meta-analysis, see Sills et al., 2016). For example, in one study (Doise & Mugny, 1978), 5- to 8-year-olds were asked to solve a spatial reasoning task that involved reconstructing a model. Initially, children worked individually, then with a partner who had either similar or different views about how to reconstruct the model, and finally individually again. Results revealed that children who interacted with partner holding opposing viewpoints exhibited stronger improvements in their individual spatial reasoning abilities compared to those working with partners sharing the same viewpoints. As a result of encouraging findings like these, teaching methods centered around disagreement have become integral components of many educational curricula (see, e.g., Driver, 2000; Kokotsaki et al., 2016; Osborne, 2010; Rapanta & Felton, 2022; Wood, 2003). Within the framework of cooperative learning, for instance, students are frequently prompted to engage in activities that challenge them to negotiate their different viewpoints (see Johnson & Johnson, 2009).

While the individual psychological benefits of disagreement are widely recognized, the specific cognitive mechanisms that lead to learning and cognitive growth remain underexplored. Piaget (1952), for example, proposed that disagreements (and other forms of cognitive conflict)

trigger mental accommodation, where children adjust or form new cognitive schemas to reconcile conflicts between their perspective and alternative views, leading to cognitive change. However, Piaget did not specify the cognitive processes involved in accommodation. In contrast, the current dissertation explores one specific potential answer to the question of how disagreement might generate learning and cognitive change: namely, that disagreement prompts children's metacognitive reflection, and leads them to ponder questions such as 'How do I know what I know?', 'How confident am I in my beliefs?', and 'Does the evidence justify my belief, or should I consider alternatives?'. The ability to consciously reflect on one's own knowledge and understanding in such ways has been widely recognized as a cornerstone of lifelong learning (see, e.g., Roderer & Roebers, 2014).

Theoretically, the idea that disagreement fosters metacognition is rooted in cultural evolutionary accounts of metacognition, which have recently proposed that human metacognitive capacities originate in, and are adapted for, conversation-based social interactions (Dutilh Novaes, 2018; 2020; Heyes, 2018; Kuhn, 2019). In the remainder of this chapter, I will first introduce cultural evolutionary accounts of metacognition in more detail. Next, I will explain why disagreement constitutes a promising candidate for a specific social context that might be driving metacognitive development. Then, I will review initial evidence suggesting that young children's metacognitive reflection benefits from disagreement, and explain which important outstanding questions are addressed by the research presented in this dissertation. Finally, I will provide an overview over the rest of this dissertation by previewing the main findings from Chapters 2 to 4 and outlining the future research directions to be discussed in Chapter 5.

1.2 Cultural evolutionary accounts of metacognition

The concept of metacognition – which is broadly defined as "cognition about cognition" or "thinking about thinking" (Flavell, 1979) – has a rich history in psychology. Under its umbrella, researchers have explored a diverse set of phenomena, ranging from philosophical reflections on the nature of knowledge and knowing (see Steup & Neta, 2024) to the ability to mentally represent and reflect upon the content of others' minds (i.e., Theory of Mind; see Kuhn, 2014). However, the most widely accepted definition of metacognition – which is also the one adopted in the current dissertation – is the capacity to consciously monitor and reflect on one's own cognitive processes (see, e.g., Kuhn, 2022; Roebers, 2017).

Traditionally, the ability to reflect on one's own knowledge has been characterized as a capacity whose main purpose it is to support individual learning and decision-making (e.g., Dunlosky & Metcalfe, 2009; Freeman et al., 2017; Schraw et al., 2006; Sloman, 1996; Flavell, 2000). Indeed, a substantial body of research shows that individuals with better explicit metacognitive abilities are generally more effective learners (Dunlosky & Metcalfe, 2009; Freeman et al., 2017; Karataş & Arpacı, 2021; Roderer & Roebers, 2014; Schraw et al., 2006). Moreover, interventions aimed at improving individuals' metacognitive monitoring (e.g., through metacognitive feedback) can improve learning outcomes among both children and adults (Michalsky et al., 2009; Metcalfe, 2009; Metcalfe & Finn, 2013; Rinne & Mazzocco, 2014).

However, recent social accounts of metacognition have argued that an additional crucial function of metacognition is to serve *social* objectives: that is, metacognition allows us to more efficiently communicate with others and make better collective decisions (e.g., Mahr & Csibra, 2018; Mercier & Sperber, 2011; Nagel, 2015; Shea et al., 2014). Supporting this idea, in the

context of a perceptual decision-making task, the combined performance of two individuals working together surpasses that of the best individual in the group, but only if the individuals are able to exchange metacognitive judgments about their confidence with each other (Bahrami et al., 2010; Fusaroli et al., 2012). In addition, individuals working together in a group are more successful at solving the classic Wason card selection task (Wason & Shapiro, 1971) than individuals working alone, and analyses of group members' communication show that they frequently exchange their explicit metacognitive states with one another during this task (see Moshman & Geil, 1998).

Among proponents of social accounts of metacognition, some emphasize exclusively the social *functions* of metacognition (i.e., its role in joint decision-making; Mercier & Sperber, 2011; Shea et al., 2014; Greven et al., 2009). Others, however, contend that the social function of metacognition makes it very plausible that its *origin* is *also* social (e.g., Dutilh Novaes, 2018; 2020; Heyes, 2018; Tomasello, 2019; Kuhn, 2019; Nagel, 2015; O'Madagain, 2019). Drawing on early sociocultural theories (e.g., Bruner, 1991; Vygotsky, 1987), these perspectives – whom I will refer to as "cultural evolutionary accounts" – suggest that metacognition is not a cognitive "instinct" acquired via genetic transmission, but a cognitive "tool" that children acquire through their conversation-based interactions with other members of their social group (see e.g., Heyes, 2018; Heyes O'Madagain & Tomasello, 2021).

By characterizing both the functions, as well as the origins of human metacognitive abilities as inherently social, cultural evolutionary accounts offer a fascinating novel theoretical perspective. However, what these accounts have not yet specified is which specific conversational contexts might constitute the key driver(s) of metacognitive development. One possibility that has been suggested is that interactions with adults might constitute the primary contexts in which children's metacognition develops (Heyes, 2018). Essentially, the idea is that the process of learning explicit metacognition akin to learning literacy, with adults explicitly teaching children how to tease apart and interpret their metacognitive feelings. Intuitively, such interpretative guidance from adults could certainly be helpful when children first learn to map their metacognitive states to labels (e.g., "this is what it feels like to be confused").

Arguably, however, a defining feature of mature metacognitive competence is its flexibility and context-specificity. That is, a good metacognitive reasoner should continuously update and adjust their understanding of their own knowledge in line with new incoming evidence and information. It is difficult to imagine how children would develop such flexible metacognitive understanding if they had to rely solely on parental instruction to interpret every unique metacognitive state that they may encounter. Thus, the current dissertation makes the case for an alternative social context as the key driver of children's metacognitive development: experiences of disagreement with peers.

1.3 Why might disagreement have a special influence on metacognitive development?

One reason for why disagreement might play a special role in children's metacognitive development is that disagreements are frequent experiences for children from early in life. Observational studies with US children show that from the age of three, children engage in disagreements with their parents or siblings on a daily basis (e.g., Dunn & Munn, 1987; Dunn & Herrera, 1997). Although engaging in conflicts with parents may be less common in other cultural contexts – particularly those characterized by steep power hierarchies between adults

and children (Hofstede, 2011) – disagreements with siblings and peers seem to constitute universal experiences for children worldwide (see, e.g., Kinoti, 2010). Thus, disagreements might naturally provide young children with plenty of opportunities to reflect on their own knowledge on a day-to-day basis.

Beyond their prevalence, disagreements may also be special in a more structural sense. It has been argued that disagreements "personify" or "embody" alternative hypotheses in the form of an actual other person, which might make it easier to recognize alternative hypotheses compared to when these are presented non-socially (Kuhn, 2019). This way of "highlighting" alternative hypotheses might be particularly important for young children, whose thinking is often characterized by egocentricity (Piaget, 1952). Moreover, prior work on modal reasoning shows that young children often struggle with representing multiple alternative hypotheses simultaneously (e.g., Leahy et al., 2020) – and it seems plausible that this difficulty may be the root cause of young children's metacognitive failures. By embodying the alternative hypothesis within another individual, disagreements may alleviate this challenge: rather than needing to represent both alternative possibilities at once (e.g., "it might be A or it might be B"), children can attribute these possibilities distinctly to two different people and represent them sequentially (e.g., "I think A, you think B"). In this way, disagreements might uniquely facilitate young children's understanding of the fact that the alternative perspectives represent different attitudes towards the same mental content (see O'Madagain & Tomasello, 2021) – an important precondition for metacognitive reflection.

Taken together, there are several reasons for why disagreement might have a special influence on young children's metacognition, and there is now increasing empirical evidence that supports this idea. I will review this emerging body of work in the following section, focusing on four key areas in which children's metacognition is typically studied: 1) information search, 2) source monitoring / reason-giving, 3) confidence ratings, and 4) belief-revision and suspension of judgment.

1.4 Prior developmental work on metacognition and disagreement

Information search

The focus of the current dissertation in on children's ability to consciously reflect on their own knowledge, which is typically measured via verbal reports (e.g., Flavell, 2000). However, rudimentary forms of metacognition can also be examined implicitly, through behavioral measures (see Goupil & Kouider, 2019 for an overview). These more automatic behavioral expressions of metacognition are presumed to be a precursor of more conscious metacognitive reflection, and they are shared with many other species, including pigeons, capuchin monkeys, and great apes (Call & Carpenter, 2001; Kepecs & Mainen, 2012; Iwasaki & Kishimoto, 2021). Thus, they are sometimes referred to as "core metacognition" (Goupil & Kouider, 2019) or "metacognitive feelings" (Goupil & Proust, 2023; Tomasello, 2024).

One particularly commonly studied metacognitive behavior is the search for information in situations of uncertainty. For example, even preverbal infants are curious about and explore objects that violate their expectations (Stahl & Feigenson, 2015), and toddlers strategically seek help from caregivers when they're unlikely to remember the location of a hidden object (e.g., because a lot of time has passed since they observed the hiding process), but not when they're likely to remember it (Goupil & Kouider, 2016). Moreover, from around age 3, children engage

in more information-seeking when confronted with incomplete versus complete (Iwasaki & Kishimoto, 2021), or confounded versus unconfounded empirical evidence (e.g., Schulz & Bonawitz, 2007; Gweon & Schulz, 2007).

Now, new research suggests that young children exhibit similarly adaptive information-seeking behaviors when experiencing disagreement (Cottrell et al., 2023; O'Madagain et al., 2022; Ronfard et al., 2018; Ronfard et al., 2021). For instance, in one study by O'Madagain and colleagues (2022), 3- and 5-year-old children observed how a reward was placed in one of two boxes and were asked to choose which box contained the reward. In one version of the experiment, participants were then presented with new evidence contradicting their initial belief about the reward's location (i.e., a physical conflict), while in another version of the experiment, participants experienced that a social partner chose the opposite box (i.e., a disagreement). Findings showed that 5-year-old children were more likely to stop and recheck the evidence both after experiencing the physical conflict and after experiencing the disagreement, compared to the respective no-conflict comparison conditions. In contrast, the 3-year-olds were not yet sensitive to the physical conflict, but did pause and recheck the evidence after experiencing disagreement. These findings are particularly interesting because they provide initial evidence for the idea that disagreement has unique benefits for children's metacognition compared to being presented with conflicting evidence in non-social ways (see also Doise et al., 1975)

Source-monitoring / reason-giving

Although implicit, behavioral forms of metacognition appear early on, decades of research on metacognitive development have shown that children often struggle with consciously reflecting and explicitly reporting on their own limited knowledge until much later in life (e.g., Bühler et al., 2023; Destan & Roebers, 2015; Flavell et al., 1970; Gopnik & Graf, 1988; Foley et al.,1983; Lipko et al., 2009; Loftus et al., 1995; Mills & Keil, 2004; van Loon et al., 2017; see Flavell, 2000; Sodian et al., 2012, for reviews). One striking example of this limitation are young children's difficulties with monitoring the sources of their own knowledge and providing reasons for their beliefs, which is considered a key metacognitive competence (Stanovich, 2011). Previous research has typically investigated the development of this ability in individual contexts; often by introducing children to some new information, and then asking them how they learned about it (e.g., Gopnik & Graf, 1988; O'Neill & Gopnik, 1991). This work has revealed that children typically fail to accurately report the sources of their knowledge up until the age of 3 or 4.

However, in these kinds of contexts, explicitly monitoring one's sources is arguably not particularly important; for instance, children frequently update their beliefs in line with source information without explicitly representing them as sources (Robinson & Whitcombe, 2003; Whitcombe & Robinson, 2000; see also Harris et al., 2018). In contrast, as highlighted above, within contexts of disagreement, source-monitoring serves a vital social-communicative function: by sharing their reasons for their respective beliefs, disagreeing individuals can determine which belief is more likely to be true, and thus make effective joint decisions (e.g., "Trust me, I saw it with my own eyes"; Mahr & Csibra, 2018; Mascaro & Sperber, 2009; Nagel, 2015; Tomasello, 2019). Does this heightened relevance lead children to exhibit earlier source-monitoring competencies than previously believed when these are investigated within contexts of disagreement?

Two recent studies suggest that this is the case (Baer et al., in prep; Mahr et al., 2021). In one study (Mahr et al., 2021) 3- and 4-year-olds learned about the contents of a container (either through direct observation or through being told what was inside by an experimenter); then, an interlocutor (who had previously been absent) asked children what was inside the box, either agreed or disagreed with children's claim, and asked them how they knew what was inside. While 3-year-olds often failed to report their sources accurately – as in prior research (Gopnik & Graf, 1988) - 4-year-olds showed a facilitative effect of disagreement: after the interlocutor had agreed with them, their performance was similar to that of 3-year-olds; however, after the interlocutor had disagreed with them, 4-year-olds were significantly more likely to report their sources accurately. Using a similar paradigm, Baer and colleagues (in prep) replicated these findings in trials where children had directly observed the contents of a container, but found no differences between disagreement and agreement in trials where children had been told about the contents. One explanation for why first-hand sources (like seeing what's inside a container) might be especially likely to benefit from disagreement is that they carry more persuasive weight than second-hand sources (like being told what's inside); thus, they might be particularly useful for resolving the disagreement and making effective joint decisions.

While reporting one's reasons for one's beliefs when explicitly prompted is an important metacognitive capacity, true metacognitive competence is arguably characterized by the ability to *spontaneously* share one's reasons to enable mutual understanding. Some prior work has shown that such spontaneous reason-giving, too, can be facilitated by disagreement (Köymen, Jurkat et al., 2020; Köymen, O'Madagain, et al., 2020; Köymen & Tomasello, 2018; Mammen et al., 2019; Mascaro et al., 2019). However, this research has been carried out exclusively with children from Western cultural contexts. This is problematic because it has been argued that Western culture and Western parenting practices tend to place a particular emphasis on reason- and source-information; more so than many other cultural contexts (e.g., Gambetta, 1998, Gunaratne, 1998; Gunaratne & Shelton, 2006; Nisbett et al., 2001). If disagreement is truly a key driver of children's metacognitive development, it should affect children's reason-giving similarly across cultures, regardless of their specific cultural experiences. The goal of the study presented in Chapter 2 of this dissertation was to test this important hypothesis by adopting a cross-cultural approach and examining whether disagreement (versus agreement) prompts children from China, Kenya, and the US to provide reasons for their beliefs.

Confidence ratings

As illustrated above, even preverbal infants demonstrate some sensitivity to (un)certainty in their behaviors (see Goupil & Kouider, 2019). However, for metacognition to fully support its crucial individual (learning) and social functions (effective joint decision-making), individuals must be able to consciously access their perceived certainty levels and express them through explicit confidence ratings (e.g., "I am 70% that there is a reward in this box."; see Heyes et al., 2021; Johnson & Fowler, 2011; Shea et al., 2014). Here, young children often exhibit a metacognitive bias: they frequently express higher confidence than is justified based on their actual likelihood of providing a correct answer (e.g., Lipko et al., 2009; Flavell, 2000). For example, they largely overestimate their ability to recall images or solve mathematical problems (Bühler et al., 2023; Destan & Roebers, 2015; Flavell et al., 1970; Lipko et al., 2009; van Loon et al., 2017), state they have understood intentionally convoluted instructions (Markman, 1977), or claim to know how complex machines, like toasters, work (Mills & Keil, 2004).

Given how important accurate confidence ratings are for both effective learning and joint decision-making, researchers have often attempted to reduce young children's overconfidence – for example, by providing them with feedback on the correspondence between the accuracy of their answers and their confidence judgments (e.g., Buehler et al., 2023; van Loon & Roebers, 2020). While such feedback-based interventions effectively reduce overconfidence in older children and adults (e.g., Callender et al., 2016), they have typically yielded little to no success in younger children (e.g., Buehler et al., 2023).

However, a recent observational study hints at a promising alternative (Viana et al., 2023): the experience of disagreement. In this study (Viana et al., 2023), pairs of 5- to 9-year-old children solved a spatial transformation task, in which they had to reconstruct a miniature village from a model. Children's spontaneous verbal exchanges during the task were analyzed to determine whether children demonstrated "epistemic humility", here defined as the "willingness to be open to another's input". Crucially, one aspect of epistemic humility was whether children expressed uncertainty about their own ideas; indeed, children of all ages frequently expressed such uncertainty when faced with their partner's conflicting ideas. Hence, these findings offer preliminary evidence that in situations of disagreement, children are both able and willing to acknowledge uncertainty, challenging their perception as little "know-it-alls" (Hagá & Olson, 2017). However, the open-ended paradigm of this study and the absence of a control condition limit the ability to draw specific conclusions regarding the causal impact of disagreement on reducing young children's overconfidence. In contrast, the studies presented in Chapter 3 of this dissertation provide a controlled experimental test of this idea, by investigating whether disagreement (versus agreement) reduces young children's overconfidence in their beliefs.

Belief-revision and suspension of judgment

In addition to recognizing that one could be wrong and reducing one's confidence in one's initial belief, an additional important indicator of metacognitive competence is to attempt to reasonably *integrate* one's own perspective with the conflicting perspective that one encounters during the disagreement (see Frances, 2014; Kuhn, 2022). Traditionally, theorists have been skeptical regarding children's capacity to integrate multiple perspectives in sophisticated manners (e.g., Kuhn et al., 2000). However, recent research suggests that in contexts of disagreement, children navigate this task surprisingly well (Miosga et al., 2020; Rakoczy et al., 2021; Schleihauf et al., 2022; Yang et al., 2023; Young et al., 2012).

For example, in one recent study (Schleihauf et al., 2022), 4- to 5-year-old children in one condition were provided with strong evidence regarding which of two boxes contained a reward (e.g., children learned that one box was heavier). When confronted with a puppet who disagreed with them about which box contained the reward, children in this condition hardly ever adopted the opposing belief of a disagreeing other when this belief was supported by a bad reason (e.g., "I think the reward is in the other box because it is my favorite color"); but they did so sometimes when the contradictory belief was supported by a good reason ("I think the reward is in the other box, because I saw it in there"). In contrast, in another condition, children's initial belief was based on a guess. These children almost always adopted the belief of the disagreeing other, regardless of whether the other's belief was supported by a good or a bad reason. Thus, children seem to reflect on their own knowledge and compare the quality of their own evidence to the evidence supporting the belief of a disagreeing other when deciding whether or not to change their minds.

However, an additional important indicator of metacognition is the ability to suspend judgment when encountering disagreements in which the evidence supporting one's own belief and that supporting the belief of the disagreeing other are equally strong (Frances, 2014). By many epistemological accounts, withholding one's judgment upon realizing that the available evidence is not as good as initially perceived constitutes one of the most meaningful markers of metacognitive reflection (Crawford, 2004; 2022; Frances, 2014; Friedman, 2013, 2017; Turri, 2012). Some prior work has investigated children's understanding suspension of judgment in third-party contexts, and found that around age five, children begin to understand that when two disagreeing agents have similarly convincing evidence, they can "both be right" (e.g., Heiphetz et al., 2013; Heiphetz et al., 2017; Yang et al., 2023). However, no prior studies had investigated whether children actually withhold their *own* judgment when encountering disagreement. To close this gap, the studies presented in Chapter 4 investigate whether children selectively revise their beliefs or suspend judgment depending on the evidence supporting their own belief versus that of a disagreeing other.

1.5 Scope of this dissertation

Taken together, the research presented in this dissertation integrates two influential theoretical perspectives concerning the role of social interaction in cognitive development. The first are constructivist learning theories, which have long posited that experiencing disagreement can foster learning and intellectual growth (e.g., Dewey, 192; Moshman, 1995; Piaget, 1952; Vygotsky, 1987). However, these theories have not focused on the cognitive processes through which disagreement might generate learning. The second are cultural evolutionary accounts of metacognition (e.g., Dutilh Novaes, 2018; 2020; Heyes, 2018; Kuhn, 2019), which have emphasized the importance of conversation-based social interaction for metacognitive development, but have yet to pinpoint which specific type of social interaction might be most effective in driving this development (e.g., adult instruction versus disagreement).

By identifying interpersonal disagreement as a specific form of social interaction that might foster young children's metacognitive reflection, the current dissertation aims to advance and integrate these two theoretical fronts. Across five studies, it tests the impact of disagreement on three key aspects of young children's metacognition (see Crawford, 2022; Kuhn, 2022): their reason-giving during joint decision-making (Chapter 2), their confidence ratings (Chapter 3), and their belief-revision and suspension of judgment (Chapter 4). All studies were preregistered, and data, analytic code, photos of study materials and the experimental protocols are publicly available on OSF.

Chapter 2[1] adopts a cross-cultural approach to examine whether disagreement prompts children from three diverse cultural backgrounds (China, Kenya, and the US) to provide reasons for their beliefs. Five- to 9-year-old dyads of children were presented with a collaborative decision-making task, in which they had to jointly choose one of two possible options to acquire a reward. Before the decision-making phase, each child in a dyad individually received evidence regarding the location of the reward. In the agreement condition, children's individually acquired evidence pointed toward the same reward location, while in the disagreement condition, children's individually acquired evidence supported opposing locations. Findings revealed that,

[1] A version of this work has been submitted for publication as Schleihauf*, Langenhoff*, Zhang, Wang, Herrmann, Köymen, Zeidler, & Engelmann (* equal contribution as first author).

across ages and cultural contexts, children provided more reasons in contexts of disagreement versus agreement. This suggests that disagreement prompts children to spontaneously reflect on and share their own thought when making collaborative decisions with others.

Chapter 3[2] asks whether experiencing disagreement reduces young children's overconfidence and leads them to develop a more accurate understanding of the limitations of their own knowledge. Across two studies, 4- to 6-year-old US children experienced either a disagreement or an agreement with a confederate about a causal mechanism after being presented with ambiguous evidence. We measured children's confidence in their belief before and after the (dis)agreement, and how long children searched for information about the correct answer. Disagreement, especially with an expert (Experiment 2), reduced overconfidence and prompted children to search longer for information, compared to agreement. Together, these findings suggest disagreement leads children to more accurately represent their own knowledge. They also suggest that implicit (i.e., information search) and explicit (i.e., confidence judgments) forms of metacognition might be affected by disagreement in similar ways.

The focus of Chapter 4[3] is on whether disagreement leads children to consider the reliability of the evidence supporting their own beliefs, and flexibly revise these beliefs or choose to suspend their judgment until they have acquired more evidence, thus consciously recognizing their uncertainty (see Crawford, 2022). Across two experiments, US children aged 4 to 9, and adults formed an initial belief, and were confronted with the belief of a disagreeing other, whose evidence was either weaker than, stronger than, or equal to the participant's own evidence. With age, participants were increasingly likely to maintain their initial belief when their own evidence was stronger, adopt the other's belief when their evidence was weaker, and suspend judgment when both had equally strong evidence. Thus, disagreement seems to prompt children to reflect on their own knowledge and adjust their beliefs according to evidence.

Taken together, the findings presented across this dissertation clearly show that disagreement prompts young children's metacognitive reflection. However, many open questions regarding the relation between disagreement and metacognition remain. Three particularly important open questions which will be discussed in detail in Chapter 5. First, what is the underlying psychological mechanisms through which disagreement might be generating its beneficial effects on metacognition, and how could it be tested? Second, the current work shows that disagreement prompts children's metacognitive reflection in the moment of the disagreement itself. However, the broader idea underlying this dissertation is that disagreement constitutes the key driver of children's metacognitive development, as a whole. What do we currently know regarding this broader claim, and which future work is needed to test it further? Third, the current work has some clear practical implications; understanding the processes through which disagreement leads to learning could help make disagreement-based teaching methods even more effective, and also extend them to younger children. In addition, it suggests possibilities for interventions aimed at fostering important citizenship skills. However, disagreement can clearly also lead to negative consequences. In the final section of Chapter 5, I will discuss which factors could potentially modulate the beneficial effects of disagreement and how they could be studied.

[2] A version of this work has been published in Child Development as Langenhoff, Srinivasan, & Engelmann (2024), and can be found at http://doi.org/10.1111/cdev.14098.

[3] A version of this work has been published in Child Development as Langenhoff, Engelmann, & Srinivasan (2023), and can be found at https://doi.org/10.1111/cdev.13838.

Chapter 2: Disagreement prompts children's reason-giving

2.1 Introduction

From fish to pigeons, and baboons to humans, species across the animal kingdom make decisions as a group (e.g., Conradt et al., 2003). Collective decisions are sometimes easy and straightforward; other times, they are hindered by conflicting preferences, interests, and beliefs. Failure to reach consensus in the face of disagreement can lead to disruptions, and may even result in a group's dissolution and eventual collapse. To maintain group cohesion and overcome conflicts of interest, many species have evolved complex decision-making procedures (e.g., Strandburg-Peshkin et al., 2015; 2018; Seyfarth et al., 2003). Humans often make group decisions similarly to other animals (e.g., Bousquet et al., 2011; Walker et al., 2017). In addition, humans display a unique capacity for collective decision-making: the ability to provide information about the sources of / reasons for their beliefs. As discussed in Chapter 1, sharing the reasons for one's beliefs constitutes a key metacognitive capacity (Stanovich, 2011) that plays a crucial role in facilitating effective joint decision-making (see Heyes et al., 2020). For peaceful and egalitarian coexistence, this capacity is indispensable, as it allows for disagreements to be settled not by the coercive force of dominance but by the 'unforced force' of the better reason (Habermas, 2001; Sen, 2003).

Rational discourse can fulfill its function of fostering consensus and circulating reliable information only if the capacity to share the reasons for one's beliefs with others develops reliably during ontogeny. Yet very little is known about the developmental origins of this important metacognitive capacity. In addition, as outlined in Chapter 1, the little research that does exist has been carried out exclusively with children from Western cultural contexts; there exists, to date, no study of reason-giving in children from other cultural backgrounds. Addressing this gap is important from a methodological perspective, as it allows us to reduce the persistent sampling bias in (Developmental) Psychology (Nielsen et al., 2017). It is also crucial from a theoretical perspective, as studying reason-giving across cultures speaks to the long-standing debate about the WEIRDness (Western, Educated, Industrialized, Rich, Democratic; Henrich et al., 2010) of reason-giving as a form of collaborative decision-making.

Some theorists view the ability to give reasons for one's beliefs as a universal and early developing part of human nature (Sen, 2003; Tomasello, 2019) – with some even arguing that the evolved function of advanced human reasoning skills is to exchange reasons (Mercier & Sperber, 2011; Mercier, 2011). Others maintain that the focus on reasons and justifications is predominantly a Western phenomenon, with historical roots in classical Greece, and suggest that other cultural contexts prioritize different ways of conflict resolution (Gambetta, 1998, Gunaratne, 1998; Gunaratne & Shelton, 2006; Nisbett et al., 2001). For example, many Eastern cultures tend to place a high emphasis on harmony (Kitayama et al., 2009). In these cultural contexts, the exchange of reasons might be perceived as too confrontational. Similarly, in societies characterized by steep power hierarchies, conflicts may be resolved more often through unilateral decisions by more powerful individuals, without need for justifications (Hofstede, 2011). To begin to address the fundamental question of the universality of reason-giving, the current study investigated how this capacity develops in children from three diverse cultural backgrounds: China, Kenya, and the United States.

What we currently know about the development of reason-giving is based exclusively on research with Western children, who start to give reasons when making joint decisions with peers when they are between 4 and 5 years of age (Domberg et al., 2019; Köymen et al., 2020; Köymen & Tomasello, 2018; Köymen & Tomasello, 2020; Mammen et al., 2019). However, preschoolers' reason-giving is still limited, and they do not always share their reasons to enable mutual understanding (e.g., Köymen & Tomasello, 2018). Over the early school years, Western children's reason-giving abilities become increasingly nuanced, such that 7-year-olds almost always share their reasons for their beliefs when this is needed to establish common ground, and even explicitly compare the quality of opposing reasons to each other (e.g., "you only heard it but I saw it with my own eyes"; see Köymen & Engelmann, 2022; Köymen et al., 2020). However, whether and how children from other cultural contexts use reasons as a tool for joint-decision making, and whether there are differences in the developmental pattern, has not been tested in previous research.

One theoretical reason for thinking that the reason-giving of Western children might differ from that of children in other cultural contexts is that Western adults have been argued to communicate with their children in specific and distinctive ways. For instance, Western adults generally tend to provide their children with more one-on-one verbal input than parents in many other cultural contexts (e.g., Cristia, 2019; Heath, 1983; Lieven, 1994), and – most relevant to the current question – often present reasons for actions and beliefs to children and encourage children's participation (e.g., Baumrindt, 1971). In contrast, adults in many other cultural contexts –- including both Kenyan (Bornstein, 2013) and Chinese (Bond, 1998; Wang, et al., 2024) contexts – tend to be more likely to set strict rules and expectations for children without providing justifications or room for negotiation (Regalado et al., 2004; Bornstein, 2013).These differences in how adults from different cultural settings tend to speak to children might manifest in qualitative and quantitative differences in children's reason-giving.

However, recent research has challenged the notion of a cross-cultural categorical distinction between parenting practices, and instead maintains that contextual and situational factors are more meaningful in explaining when adults present children with reasons and justifications (Prevoo & Tamis-LeMonda, 2017; Smetana, 2017). In addition, and more importantly, parent-child interactions might not represent the main socio-interactional context in which children's skills of reason-giving develop. Piaget (1952) argued that children's reasoning abilities may benefit in particular ways when cognitively working through disagreements with epistemic peers (as in child-child interactions) rather than epistemic authorities (as in adult-child interactions). This is because in cases of disagreement with a seemingly all-knowing adult, children might simply defer to the adult's opinion. In contrast, when disagreeing with a peer, children might be more inclined to view conflicting claims as equally valid, which may prompt the exchange of reasons (Köymen & Tomasello, 2020; Kruger, 1992). Supporting this idea, children from the US have been shown to engage in more sophisticated moral reasoning with same-aged peers than in discussion with their mothers (Kruger & Tomasello, 1986; Mammen et al., 2019). This perspective raises the possibility that across cultures, children might be most likely to give reasons when they encounter disagreements with peers – our focus in the current study.

To investigate the development of reason-giving across diverse cultural contexts, we tested 180 dyads of 5- to 9-year-old children in China, Kenya, and the US in a joint decision-making task. Dyads of same-aged peers had to collaboratively decide which of two boxes they

wanted to open to obtain a reward (see Figure 1). The procedure consisted of two phases. In Phase 1 –- the evidence-gathering phase –- children were separately exposed to evidence regarding the reward's location. One child received perceptual evidence (they saw where the reward was hidden) whereas the other child received testimony-based evidence (they were told by an experimenter where the reward was hidden). Importantly, we manipulated whether the evidence that children gathered individually supported the same or different conclusions. In the disagreement condition, children were presented with conflicting evidence (for example, one child watched how the rewards were hidden in a blue box, while the other child was told that the rewards were in a red box). In the agreement condition, children received evidence supporting the same conclusion (for example, one child watched how the rewards were hidden in a blue box, while the other child was told that the rewards were in the blue box). In Phase 2 –- the joint deliberation phase – children came together to jointly decide where the reward was hidden (and which box to open). Children participated in two trials of the same condition (disagreement or agreement), with individual children observing the same kind of evidence (perceptual or testimony) across trials. We also conducted, before the test phase, a dominance test to determine which of the two children in a given dyad was the more dominant one. We included this test to evaluate whether children – in addition to using a cooperative strategy to resolve disagreements, namely reason-giving – also used a competitive strategy, namely dominance.

We tested whether children exchange reasons to make joint-decisions, whether they selectively provide reasons in contexts of disagreement, and whether the development of reason-giving show uniformity across diverse cultural contexts. We also tested whether children's reason-giving would increase with age (Köymen et al., 2020), and whether children would give fewer reasons after establishing common ground, i.e., after the second compared to the first test trial (Köymen et al., 2016). Additionally, we examined whether children from different cultures and age groups differed regarding which box they decided to open (i.e., the box supported by perceptual vs. testimonial evidence). Finally, in an exploratory analysis, we tested whether children's decision-making was influenced by dominance. All experimental procedures, hypotheses, and analyses were preregistered on OSF.

Figure 1

Example of study set-up in the three cultural contexts.

2.2 Experiment

Methods

Participants

In total, data from 360 5-to-9-year-old children from China, Kenya, and the US, who were paired into N = 180 dyads (60 dyads per culture), were included in the final sample of the study. The required sample size was determined with a power simulation (code and results are available on OSF), expecting a significant effect of the factor condition. A sample size of 180 dyads led to an average power of (1- beta = 0.73). We tested 7 additional dyads in China and 2 in the US. These additional dyads were excluded due to experimental errors (1 in China, 2 in the US), because children misremembered which box was supported by evidence (3 in China), or because children peeked into the boxes before making their decision (3 in China).

Chinese children came from peripheral areas of Beijing, belonging predominantly to the Han ethnic group, with a significant number coming from migrant families. These children typically resided with their parents and no more than one sibling \cite{binah2014discourses}. The majority of their parents had attained a high school education as their highest level of formal schooling. Most children came from low socio-economic backgrounds with many parents engaged in informal employment, such as selling groceries or washing dishes. In China, Confucian principles advocate for enduring parent-child bonds and stress the significance of familial harmony and obedience (Binah-Pollak, 2014; Sun &Ruder, 2016; Xu & Hamamura, 2014). Within the communities studied, parenting approaches are deeply rooted in these traditional values, characterized by a tendency towards punitive measures and strictness, coupled with a lesser degree of warmth and understanding, especially when compared to urban parenting norms (Wang et al., 2024).

Kenyan children in our sample came from rural, low socioeconomic status (SES) households in Nanyuki, Laikipia County, and were mainly of Kikuyu ethnicity. Traditionally, the Kikuyu have been small-scale farmers and practicing animal husbandry for their subsistence, but in recent years, trade and wage work have become increasingly important. Although children are highly valued, many families nowadays restrict their number for economic reasons (Price, 1996). Despite the trend towards smaller families, children typically grow up surrounded by siblings, cousins, and friends of various ages and have few or no possessions of their own. Many children attend the local nursery school from about 4 years of age, and almost all of them go to school once they are 5-6 years old. Outside school, children typically help with various tasks around the house, attend to younger siblings, or look after animals (Whiting, 1996). Relationships between adults and children are often strictly hierarchical (Kinoti, 2010)

US children were from the California Bay Area and had mixed racial and socioeconomic backgrounds. Children's racial and ethnical make-up was representative of urban areas in Northern California, which is approximately 50% White, 20% Asian, 10% Black, 10% Hispanic or Latinx, and 6% from two or more races. Children in the US typically grow up with their parents and one or two siblings. Other family members often live in other parts of the country and are not part of the children's daily lives. Parents often emphasize their children's psychological autonomy from an early age and children receive high levels of direct, child-centered pedagogy (Keller, 2022). Children commonly grow up with many toys and possessions. From an early age, the majority of children are cared for in nurseries and kindergartens and by six years of age, children attend primary school (National Center for Education Statistics, 2017).

Ethical Statement and Data Availability

The study was conducted in accordance with the Declaration of Helsinki and the ethical guidelines of the Ethical Review Board of the University of California, Berkeley. In China and the US, children's caregivers gave informed consent prior to their child taking part in the study. In Kenya, we obtained permission to conduct the research from the National Commission for Science, Technology, and Innovation and the local school boards. Informed consent for children's study participation was given by the schools' head-teacher. Anonymized behavioral data and reliability coding, as well as the preregistration of the study have been deposited in the Open Science Framework (https://osf.io/bmy7s/?view_only=1a90ffe5017a49dbb14075a9dfdb4179).

Procedure

Children were paired into dyads matched by gender and age, with a maximum age difference of one year between them. The study followed a between-subject design, with half of the dyads participating in the disagreement condition, the other half in the agreement condition. In both conditions, children first participated in a dominance test. Our assessment of dominance was defined in terms of the capacity to exert control over resources (Hawley, 1999). To measure which child in a dyad was the more dominant one, we placed an attractive toy between the two children and left them alone. In line with prior work (Grueneisen & Tomasello, 2022), the child who took the toy first was identified as the more dominant child of the dyad. Following the dominance test, children participated in two training trials, that were identical across conditions, then they participated in two test trials, that differed according to condition. In all trials, children were presented with a decision-making task. Dyads were presented with a pair of boxes and were informed that one of the boxes held two rewards, while the other was empty. They were instructed that they could, as a team, only choose one of the boxes; if they opened the correct box, both children would get a reward, but if they opened the wrong box, none of them would get a reward. We presented a different pair of boxes in each trial. The boxes belonging to one pair, always had different colors, so that it was easy for children to refer to the individual boxes. The study was run by two experimenters; one experimenter led the study (and provided testimonial evidence in test trials), the other experimenter took care of preparing the boxes (and provided perceptual evidence in test trials). Children were given a maximum of two minutes for their collaborative decisions about which box to open, but they could also call the experimenters back once they were ready. While children discussed, both experimenters stood far enough away from the children (or if possible, left the room), so that children did not feel like they were being overheard. When the experimenters returned, children were given a small rock to place on top of and indicate their chosen box. This method ensured that only one box was chosen, even in the absence of a prior mutual agreement between the children.

Training trials. In the training trials, children received no evidence regarding which box held the rewards, requiring them to guess. Both boxes (i.e., the one chosen in the first and the one chosen in the second training trial) were opened after the end of the second training trial. For all children, the choice of the first training trial led to a successful outcome (i.e., both boxes had been endowed with rewards), while the choice of the second trial led to failure (neither box had been endowed with rewards). Thus, children were left with the impression that they made the right choice in the first training trial and the wrong choice in the second training trial.

Test trials. In the test trials, each experimenter asked one of the children to follow them into opposite corners of the room, where children individually received evidence regarding which box held the rewards. In the disagreement condition, children received contradicting evidence about the reward's location: One child witnessed one experimenter placing the rewards in one box (perceptual evidence), while the second child was told by the other experimenter that the reward was in the other box (testimonial evidence). In the agreement condition, children received matching evidence: One child witnessed one experimenter placing the rewards in one box (perceptual evidence), while the second child was told by the other experimenter that the reward was in the same box (testimonial evidence). To ensure that children were not confusing color words, the testimonial experimenter additionally pointed at photograph of the respective box while providing testimonial evidence. As in training trials, the box selected in the first test trial was set aside. Only after the second test trial, both boxes were opened. In the test trials, all boxes held rewards. As a result, children were left with the impression that they had correctly chosen in both instances, and they received small toys as rewards from the boxes.

Data Recording and Coding

The full study procedure was video-recorded, so that we could transcribe and analyze children's conversations leading to their decisions in detail. Children's conversations during the test trials were transcribed and, where necessary, translated by native speakers into English. Using the English translations, we coded whether the dyads mentioned a perceptual reason and/or a testimonial reason. If they mentioned only the perceptual or the testimonial reason, they received a score of 1, if they mentioned both types of reasons, they received a score of 2, if they mentioned no reason, they received a score of 0. Reasons other than the cued perceptual or testimonial reason were not counted (e.g., "I want red because it is my favorite color"). A second coder coded 25\% of the data for reliability purposes. Reliability was very high with \(\kappa = 0.93\). To analyze children's box choices, we coded whether the dyads chose to open the box that had been supported by the perceptual or the testimonial reason. For this variable, the interrater reliability was perfect with $\kappa = 1$. To analyze whether power dynamics influence children's choices, we additionally coded whether dyads chose the box for which the dominant child had received evidence. Here, we reached an interrater reliability of $\kappa = 0.8$, which indicates strong agreement. Description of additional coding for additional analyses can be found in the Supplementary Material.

Data analyses

Statistical analyzes were performed in R Version 4.2.3 (2 Core Team, 2021) using Generalized Linear Modeling (Baayen et al., 2008) with the package lme4 and the function glmer (Bates et al., 2015). We fitted three models.

With the first model, we analyzed children's reason-giving tendencies. The response variable reason-giving had a lower (0 reason given) and upper bound (2 reasons given). Such variables can be transformed into a binary variable by scoring each given reason (perceptual or testimonial) as 1 and a reason that was not mentioned as 0. Thus, we analyzed these data with a binary response term and a binomial error structure. We generated a full model that contained a two-column response matrix, with the first column indicating reasons given per dyad and trial and the second column indicating reasons not given per dyad and trial. For example, if a dyad mentioned one reason (e.g., the testimonial reason), but not the other (e.g., the perceptual

reason), both columns contained a 1. As fixed effects, we included an interaction of condition, age (z-standardized), and culture, as well as the main effects of test trial and gender. Dyad identity was included as a random effect with the random slope of test trial (dummy coded and centered). The correlation between random slope and random intercept were removed because it could not be determined. We further fitted intermediate models that lacked interaction terms, as well as a null model that only contained the intercept and the random effects. See Supplementary Material for a detailed description and model equations. To avoid an increased type 1 error risk due to multiple testing, we first tested the overall effect of all test predictors by comparing the full model with a null model comprising only the random effects. This full model explained the data better than the null model ($\chi^2(13) = 44.84$, $p < .001$). To determine the effects of each predictor alone, we used R's drop1 function with likelihood ratio tests on reduced model that contained the term of interest. We depict estimates and their 95% CIs (retrieved from 1000 parametric bootstraps) of all significant effects in the figures.

With the second model, we analyzed children's box choices. The testimonial and perceptual evidence pointed to distinct boxes only in the disagreement condition. Thus, only the data of the disagreement condition was analyzed. Since the response variable box choice was binary (testimonial box or perceptual box), we fitted a model with a binomial error structure. As fixed effects, we included an interaction age (z-standardized) and culture. Dyad identity was included as a random effect with the random slope of test trial (dummy coded and centered). As in the prior model, the correlation between random slope and random intercept could not be identified, thus, we removed it from the model. We further fitted a main effect model that lacked interaction term, as well as a null model that only contained the intercept and the random effects. Model comparison revealed that the full model did contribute to the explanation of the data compared to the null model ($\chi^2(5) = 10.99$, $p = .052$). Thus, we continued to determine the effects of the individual predictors using the drop1 function.

With the third model, we analyzed how often dyads chose the box the more dominant individual had received evidence for. Again, only data of the disagreement condition was analyzed and a model with a binomial error structure was fitted. We included the fixed effects of age (z-standardized) and culture and their interaction. We also included the random intercept of dyad identity and the random slope of test trial (dummy coded and centered). The correlation between random slope and random intercept was again removed from the model. We again fitted an additional main effects model and a null model. Model comparison revealed that the full model did not explain the data better than the null model ($\chi^2(5) = 4.40$, $p = .439$). Nevertheless, we determined the effects of the individual predictors using the drop1 function. Further information as well as additional analyses can be found in the Supplementary Material. R Code to reproduce analyses and figures is available at OSF.

Results

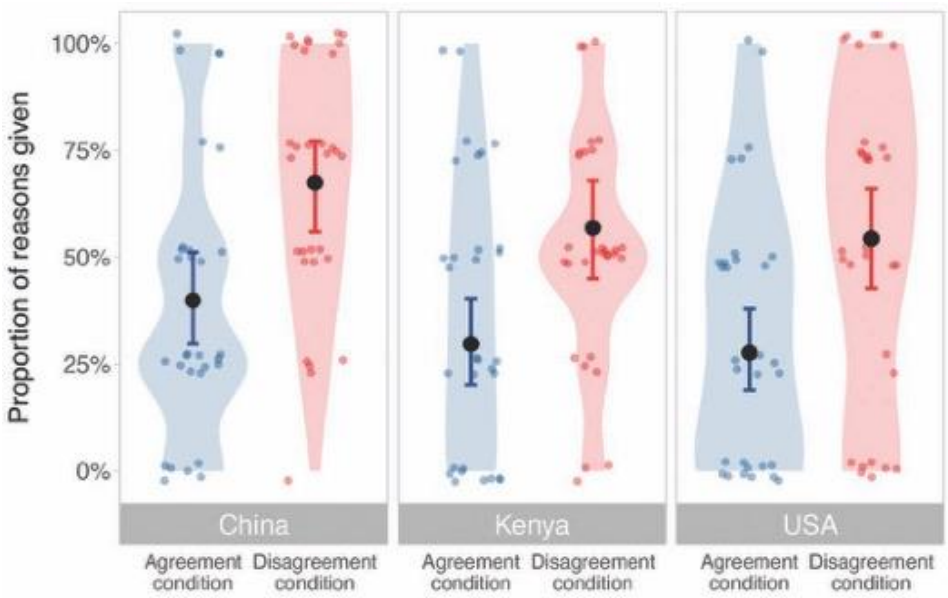
Figure 2 depicts estimates with their 95% confidence intervals for children's probability to give reasons. It shows that children across all three societies were more likely to give reasons in the disagreement compared to the agreement condition. These condition differences were statistically reliable: The effect for the predictor condition reached significance ($\chi^2(1) = 21.98$, $p < .001$), while the effect for the interaction between culture and condition did not ($\chi^2(1) = 1.41$, $p = .439$). To ensure the reliability of our results, we ran an additional non-parametric Mann Whitney U Test for each culture. This test, other than the GLMM, does not have any prior

assumptions about the data. In line with our preregistered analysis, we found significant differences between the agreement and the disagreement condition for all three cultures (China: $W = 1111.5$, $p < .001$; Kenya: $W = 1418$, $p = .033$; USA: $W = 1299$, $p = .005$).

Figure 3a shows that children's tendency to exchange reasons increased with age. This finding is supported by a significant main effect of age ($\chi^2(1) = 5.25$, $p = .022$). In Figure 3b, we present the estimates of children's probability to exchange reasons as a function of trial number. This shows that children gave more reasons in the first trial compared to the second trial of the decision-making game. This difference is statistically significant ($\chi^2(1) = 7.14$, $p = .008$).

Figure 2

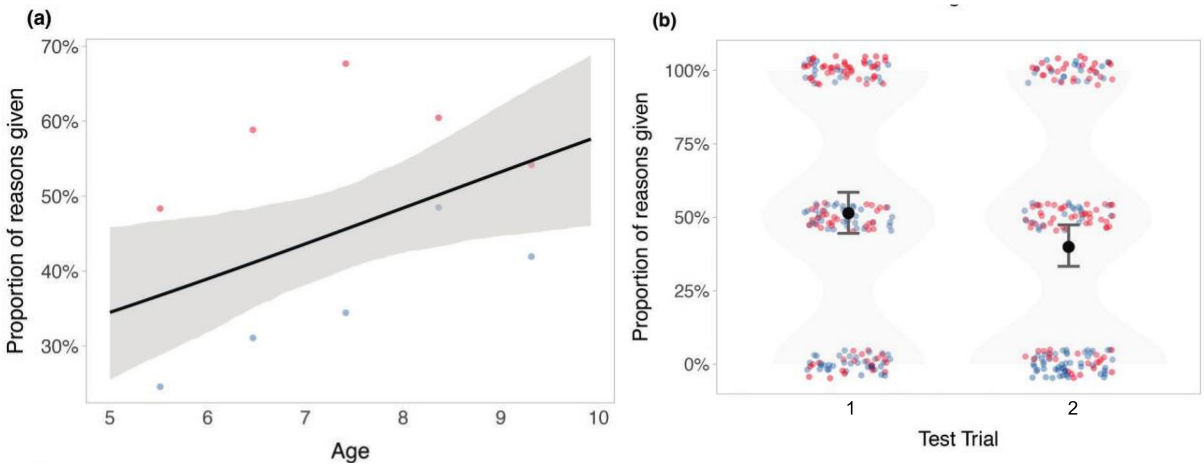
Estimates and confidence intervals of children's reason-giving as a function of the significant main effect of condition across cultures.



Note. The small dots represent the proportion of reasons given across test trials per dyad. If dyads gave 100% of reasons (4 reasons), they provided a testimonial and a perceptual reason in both test trials. The filled circle on the error bar shows the model prediction. The error bars show the bootstrapped 95% confidence intervals of a generalized linear mixed model including all main effects (the predictor variables that are not depicted were centered.

Figure 3

Estimates and confidence intervals of children's reason-giving as a function of the significant main effects of (a) age and (b) trial.

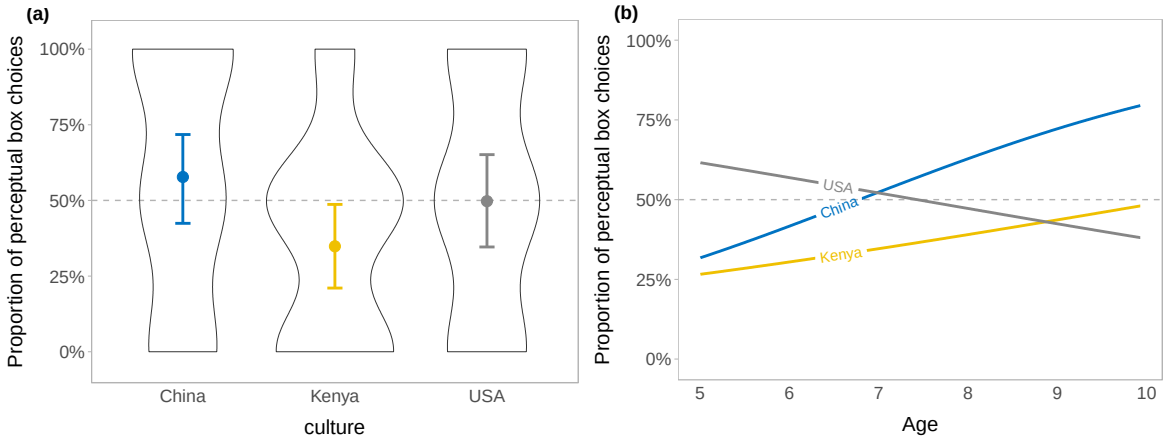


Note. In (a) the small dots represent the proportion of reasons given averaged for the dyads age, in (b) they represent the proportion of reason-giving averaged across trials per dyad. The black line (in a) and the black dots (in b) represent the model estimates. The ribbon (in a) and the error bars (in b) show the bootstrapped 95% confidence intervals of a generalized linear mixed model with all predictor variables centered except for the ones depicted.

To investigate whether dyads of varying ages and cultural backgrounds demonstrate distinct preferences for choosing between boxes supported by perceptual versus testimonial evidence, we analyzed only the data from the disagreement condition. The estimates and bootstrapped confidence intervals in Figure 4a demonstrate the probability that children choose the perceptual box for each culture. The confidence intervals of China and the United States include chance level, which indicate that children did not show a preference for any of the boxes in these countries. The confidence interval for the Kenyan children is slightly below chance level, thus, Kenyan children had a slight preference to choose the box supported with a testimonial reason. However, the estimates for each country did not significantly differ from each other ($\chi^2(2) = 4.41$, $p = .110$). Figure 4b shows the models predictions for children's box choices as a function of age and culture. Children in Kenya and China demonstrate a preference for the box supported by a perceptual reason as they get older, while there is a reverse trend observed for children from the United States. However, this trend does not reach statistical significance ($\chi^2(2) = 4.58$, $p = .101$). Overall, dyads from varied cultural backgrounds only exhibit minimal differences in their preferences to follow perceptual or testimony-based evidence.

Figure 4

Estimates and confidence intervals of dyad's choices of the box that was endorsed by perceptual evidence as a function of (a) culture only and (b) culture and age. These effects did not reach significance.



Note. The small dots represent the proportion of perceptual box choices across test trials per dyad. The lines in (a) and the big dots in (b) represent the model estimates. The ribbons and error bars show the bootstrapped 95% confidence interval of a generalized linear mixed model with all predictor variables centered except for the ones depicted.

Lastly, we analyzed whether dyads in the disagreement condition preferably selected the box the more dominant child in the dyad had received evidence for. This analysis revealed that dyads' decision-making was not driven by the dominant individual. Neither the interaction between culture and age ($\chi^2(2) = .70$, $p = .705$), nor the main effects of culture ($\chi^2(2) = 1.76$, $p = .417$) and age ($\chi^2(1) = 2.22$, $p = .136$) influenced whether the dyads followed the evidence presented to dominant individual. A detailed description of all analyses, including detailed model equations and statistical assumption checks, are provided in the Supplementary Material.

2.3 Discussion

We found that children from three diverse societies exchange reasons to resolve disagreements. Reason-giving was spontaneous: Although children were in no way trained or prompted to provide reasons during our experimental procedure, they showed an intuitive metacognitive understanding of the relevance of providing one's reason to make collective decisions and build consensus through the mutual expression of reasons. Reason-giving was also flexible: Children did not simply justify their claims independent of context. Instead, they selectively employed reasons as metacognitive tools to resolve disagreements (and gave more reasons in contexts of disagreement relative to agreement). We did not find an effect of dominance: Children showed at most a weak tendency to use dominance to settle differences of opinion. These findings suggest that by age 5, children possess key capacities for engaging in rational discourse and for resolving disagreements cooperatively – by exchanging reasons – rather than competitively – by using force or dominance.

The pattern of results was consistent across three diverse cultural contexts, China, Kenya and the US. This is remarkable because children in these contexts grow up exposed to different epistemic practices. While children in the US sample more often interact with adults who encourage joint reasoning through a focus on giving and asking for justifications (Baumrindt, 1971), children in our Chinese and Kenyan sample are more commonly surrounded by adults

who focus on deference and discourage joint reasoning (Bornstein, 2013; Wang et al., 2024). In addition, the three societies also differ on various social dimensions that have been singled out as relevant to reason-giving, i.e., the extent to which they emphasize social harmony (Kitayama et al., 2009) and the extent to which they have a more hierarchical versus egalitarian structure (Hofstede, 2011). The fact that we nevertheless observed very similar skills of reason-giving provides support for the hypothesis that the exchange of reasons in rational discourse is a cross-culturally recurrent – and potentially universal – form of resolving conflicts of interests and making collaborative decisions (Mercier & Sperber, 2011; 2017; Mercier, 2011; Sen, 2003, Pinker, 2011). Our findings are also in line with theories which argue that peer-to-peer interactions are particularly fruitful contexts for the development of children's reasoning (Köymen & Tomasello, 2020; Kruger, 1992; Piaget, 1952). While we detected no cross-cultural differences in the current study, it is possible that such differences emerge when we observe children's reasoning in different contexts, for example when they reason with authorities. One possibility that children in the US would continue to provide reasons when making joint decisions with adults, whereas children from China and Kenya might be less likely to do so. How different conversation partners influence and prompt children's reason-giving represents an exciting domain for future research.

The current experimental setup was closely modelled after how public discourse commonly happens in the real world, where individual exploration of different information sources is often followed by joint deliberation. Since our main point of interest was reason-giving, we exposed both children to strong evidence during the evidence-gathering phase (to increase the likelihood that both children would form a strong belief and express it). One child received direct perceptual evidence – they saw where the reward was hidden –- and the other child received testimony from an authority – the experimenter told them where the reward was hidden. This provided us with the ideal setup to observe children's reason-giving. However, there was no clearly stronger reason during the joint deliberation phase and no obvious solution. Thus, our study set-up was less ideally suited to test whether children also display the other main skill of rational discourse, besides reason-giving, namely reason-responsiveness: the ability and willingness to change one's beliefs when presented with better reasons for alternative views (see Kuhn & Crowell, 2013). There is evidence that children from various cultural contexts evaluate and respond appropriately to reasons communicated by others. For example, children selectively revise their initial belief when they are provided with a better reason but maintain their initial belief when they have the better reason (Langenhoff et al., 2023; see Chapter 4; Castelain et al.., 2016; Schleihauf et al., 2022; Schleihauf et al., 2023; Harris, 2012; Harris et al., 2018). In the current study, we observed some cross-cultural differences in children's preference for the perceptual versus the testimonial reason (with Kenyan children, but not children from China or the US, endorsing it at above chance levels). In future work, it would be interesting to more directly investigate whether the ranking of the strength of different reasons varies across cultural contexts.

Previous research demonstrates that even young children show remarkable skills of collaboration (Brownell et al., 1991; Hepach et al., 2018; Warneken & Tomasello, 2013). This research often focuses on physical acts of collaboration, for example when children skillfully move an object together that they could not move on their own (Hamann et al., 2011). Here, we have shown that children also engage in epistemic collaboration. When they hold conflicting beliefs, children cooperatively work through the resulting disagreements by exchanging reasons

rather than by using force. Offering someone a reason contrasts starkly with ordering or forcing them to believe something. It is a respectful and egalitarian form of address compared to the hierarchical address of command (Engelmann & Tomasello, 2019). Our results suggest that in contexts of disagreement, children from diverse societies and from at least the age of 5, reflect on their own thought their reasons for their beliefs when making joint decisions. Adults can learn a thing or two from them.

Chapter 3: Disagreement reduces children's overconfidence

3.1 Introduction

We all seem to agree that, increasingly, we do not agree on very much. Around the world, polarization is intensifying: people disagree even on matters that would seem to be indisputable, such as the reality of climate change and the safety and efficacy of vaccines (Silver et al., 2021). The proliferation of disagreement is a concern, as it is associated with heightened political extremism and violence (e.g., Iyengar et al., 2019), and reduced mental health and psychological well-being (e.g., Coleman et al., 2014). Yet, from another perspective, disagreement can also have positive consequences. Thinkers such as Gandhi and Marx have highlighted conflict as an important predecessor of societal change (Roy, 1984), and, on an individual psychological level, disagreement may prompt people to reason in more sophisticated ways (Lackey, 2010; Mercier & Sperber, 2011). The present studies explore whether certain types of disagreement can benefit cognitive development – and specifically, metacognition – by (1) reducing young children's overconfidence in their beliefs and (2) increasing children's motivation to acquire the correct answer.

As discussed in Chapter 1, the current work is grounded in social accounts of reasoning, which postulate that human higher reasoning skills, including metacognition, are best understood as fundamentally social skills, shaped by an individual's engagement in interpersonal discourse (Dutilh Novaes, 2018; Heyes et al., 2020; Kuhn, 2019; Mercier & Sperber, 2011; O'Madagain & Tomasello, 2021). While these theories generally focus on interpersonal discourse as a whole, we propose here that disagreements might be particularly fruitful context for improving children's metacognitive awareness. Specifically, disagreement might highlight diverging perspectives, and thereby lead children to metacognitively reflect on their own knowledge and understanding (see Christensen, 2009; O'Madagain & Tomasello, 2021; Lackey, 2010; Piaget, 1952). One particularly intuitive consequence of such metacognitive reflection might be that children decrease their confidence in their initial belief and increase their interest in alternative beliefs.

Although the idea that disagreement reduces overconfidence and promotes the exploration of alternative beliefs has intuitive appeal, it has to our knowledge not been tested empirically. The present studies fill this gap by investigating the effects of disagreement on overconfidence and exploration in 4- to 6-year-olds. We focused on this age group because although overconfidence persists across the lifespan (e.g., Kruger & Dunning, 1999), children of this age are particularly prone to expressing unwarranted confidence, for example, by claiming to be "really sure" even about things that they cannot know, such as what the color "byzantium" is (Hagá & Olson, 2017a; see also Lipko et al., 2009; van Loon et al., 2017). The potential effects of disagreement in reducing overconfidence could thus be especially valuable for this age group. Moreover, this is a developmental period during which children's explicit theory of mind reasoning abilities become more sophisticated (e.g., Wellman & Liu, 2004), and these developmental changes could be related to how children understand and potentially benefit from disagreements.

Understanding the effects of disagreement on overconfidence is not only important for identifying the processes that influence cognitive development, but could also play a role in informing interventions toward supporting intellectual humility (Leary et al., 2017) and learning (see Baer & Kidd, 2022; Roebers, 2017). Although possessing inflated confidence in one's be-

liefs may be adaptive for young learners in some cases (Hagá & Olson, 2017a), the ability to accurately monitor one's confidence supports important self-regulatory pro- cesses related to learning, including allocating sufficient study time and asking for help when needed (Destan et al., 2014; Hembacher & Ghetti, 2013). Indeed, over- confidence is negatively associated with learning in both older children (e.g., Freeman et al., 2017) and adults (e.g., Dunlosky & Rawson, 2012).

Although helping young children develop a more realistic understanding of their intellectual limitations may support learning across the lifespan (see Buehler et al., 2023; Lipko et al., 2009), previous attempts to reduce overconfidence in young children have led at best to minor reductions (e.g., Buehler et al., 2023; Lipko et al., 2009; van Loon et al., 2017; van Loon & Roebers, 2020). In these studies, children were typically provided with feedback on whether their answers were correct or incorrect (i.e., performance feedback), or on the correspondence between the accuracy of their answers and their confidence judgments (i.e., metacognitive feedback; see, e.g., Buehler et al., 2023; van Loon & Roebers, 2020). One explanation for the relatively weak effects of these feedback-based interventions may be that they typically focus on the child as an individual learner, rather than on social interactions that naturally highlight the presence of alternative perspectives, as in the case of disagreement—which we focus on here.

An additional goal of the current work was to investigate whether young children's overconfidence is influenced by who they disagree with. Extensive prior research has demonstrated that when children observe disagreements among others, they pay attention to the characteristics of the disagreeing individuals and preferentially adopt the beliefs of more reliable or knowledgeable individuals over those of less reliable or knowledgeable individuals (for an overview, see Harris et al., 2018). For example, when children observe a disagreement between an expert and a novice, children are more inclined to learn from the expert than the novice (Koenig & Jaswal, 2011). Here, we were interested to see whether expertise would influence children similarly if they were involved in the disagreement themselves. Specifically, we predicted that when children experienced a disagreement with a naïve confederate (Experiment 1), they would re- duce their overconfidence in their original belief only marginally, because there would be no indication that the confederate had more knowledge about the question under discussion than the child. In contrast, we expected that when children experienced a disagreement with an expert who had prior knowledge about the relevant issue (Experiment 2), children would reduce their overconfidence more significantly in light of the expert's greater prior knowledge.

Across two preregistered studies, we investigated whether disagreement reduces overconfidence and prompts exploration in four- to six-year-old children. Children and an adult confederate were exposed to ambiguous evidence about which toys make a machine play music (Figure 5). Children were then asked which toys made the machine play music and we probed their degree of confidence in their belief. Then, depending on condition, the confederate either agreed or disagreed with the child. We measured (i) changes in children's confidence in their initial belief and (ii) how long children spent searching for information about the correct answer (see Bonawitz et al., 2011; Gweon et al., 2014; Shneidman et al., 2016). In Experiment 1, children and the adult confederate had the same limited knowledge about the toys. In contrast, in Experiment 2, the adult confederate was introduced as an “expert” who knew a lot about the toys.

3.2 Experiment 1

Methods

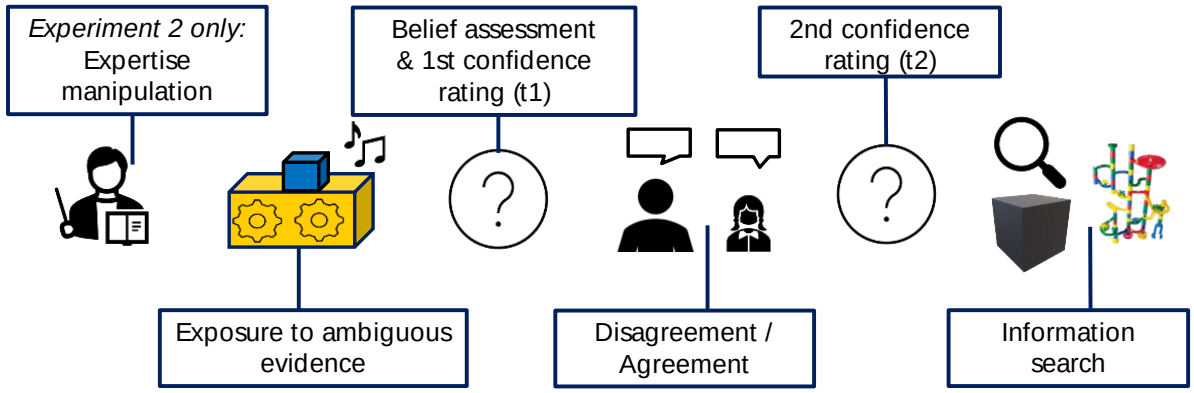
All data, analytic code, photos of study materials and the experimental protocol are publicly available on OSF and can be accessed at https://osf.io/xvkb3/?view_only=04f9b9d8cf6d4f66b26b885d6c09ad33. Both studies were preregistered on AsPredicted (Experiment 1: https:// aspredicted.org?RBC_EOX; Experiment 2: https://aspredicted.org/5YP_KYR). Information regarding power analyses and exploratory measures and analyses are stated in the preregistrations and can be found in the Supplementary Material.

Participants

Based on a power analysis (see Supplementary Material), we tested 68 children (M_{age} = 5.43 years, SD_{age} = 0.81, 36 girls, 32 boys): 33 in the disagreement condition and 35 in the agreement condition. Children were from families with mixed socio-economic status (SES) backgrounds. Forty percent of children were white, 18% had multiple races or ethnicities, 12% were Asian, and 1% African American and Hispanic, respectively (28% of parents did not provide information about their child's race or ethnicity). Seven children were excluded in line with our preregistered exclusion criteria due to not providing a confidence judgment (N = 3), experimenter errors (N = 2), the machine malfunctioning (N = 1), or sibling interference (N = 1). Information search data were not obtained for two children due to sibling interference, so the analysis of children's information search is restricted to data from the remaining 66 children. Data for Experiment 1 were collected between January 2020 and November 2021. Children were tested either in lab, in their preschool or school, at children's museums, or at a local zoo in the San Francisco Bay Area in the United States. Children only participated if parents had provided written consent and children had provided verbal assent. Upon completion of the experiment, children received a sticker as a reward for their participation. The experiment was approved by the institutional review board of the University of California, Berkeley.

Figure 5

Overview of the procedures of Experiments 1 and 2.



Procedure

The procedure consisted of five main stages: (1) introduction to the game and exposure to ambiguous evidence, (2) belief assessment and initial confidence rating (t1), (3) disagreement/agreement phase, (4) second confidence rating (t2), and (5) the information search phase. For an overview of the procedure, see Figure 5.

Introduction and exposure to ambiguous evidence. At the beginning of the experiment, the experimenter briefly introduced the confederate (played by one of multiple research assistants, who were all adults), by say- ing "This is (Name), (s)he will play with us today." After a short warm-up with a puzzle game, the experimenter announced the start of a new game. She brought out a box with her "toys," shook it, so that the participant and the confederate could hear (but not see) the toys inside, and pulled four toys out of the box, highlighting either the toys' shapes (triangle, square) or their colors (blue, green). For example, she pulled out a blue star and said, "some of my toys are blue toys, like this one." The experimenter explained that the goal of the game was to decide which of her toys were blickets, stating that only some toys were blickets, and others were not, and that the participant and the confederate could use her machine to figure this out. She explained that the machine would play music when a blicket was placed on it and would not play music when a toy that was not a blicket was placed on it. During the experimenter's explanations, the confederate listened attentively, to indicate that these rules were new for them, and they had no prior experience with the game. The experimenter then pretended to pull one toy out of her box at random (in fact, this toy was always a blue square) and handed it to the participant to "see if it would make the machine play music." Once the participant set the toy on the machine, the experimenter secretly pressed a remote control, which activated a bell system inside the box and caused it to play music. Next, the confederate got to place one toy on the machine. This toy also "happened" to be a blue square and also activated the machine. Upon hearing the machine's music, the confederate exclaimed with an excited and surprised voice: "Oh, this toy is also a blicket!".

Belief assessment and initial confidence rating (t1). The experimenter then asked the participant: "So, [child], which toys do you think are blickets? Do you think square toys are blickets? Or do you think blue toys are blickets?" (order counterbalanced). Once the participant had stated their belief about which toys were blickets, the experimenter assessed their confidence in their belief by asking "How sure are you that [square/blue] toys are blickets? Are you sure or not sure?", followed by "are you really (not) sure or just a little (not) sure?". Through this procedure, modeled after Hagá and Olson (2017), we obtained confidence ratings on a 4-point Likert scale, ranging from "1" (really not sure) to "4" (really sure).

Disagreement/agreement. Next, the experimenter asked the confederate which toys they thought were blickets. In the disagreement condi- tion, the confederate opposed the participant's belief. For example, if the participant had said that they thought square toys were blickets, the confederate said, "I don't think square toys are blickets. I think blue toys are blickets." In the agreement condition, the confederate confirmed the participant's belief. Then, the confederate asked the participant to justify their belief (e.g., "Why do you think square toys are blickets?") and justified their own belief by saying, for example, "Okay, well, I think blue toys are blickets because we put two blue toys on the machine and they made the machine play music." After that, the confederate looked at their watch, said they had to leave, and left the room.

Second confidence rating (t2). After the confederate departed, the experimenter re- minded the participant of their own and the confederate's belief, and assessed the participant's confidence in their own belief for a second time, providing the same response options as described above.

Information search. Next, the experimenter brought out another, previously hidden box and shook it, so that the participant could hear that it contained a toy. The experimenter told the participant that the person who made the machine had hidden a toy in this box that was "definitely and for sure" a blicket. While placing the box a few feet away from the participant, the experimenter said that if the participant really wanted to find out which toys were blickets, they could go ahead and open the box. The experimenter then brought out a marble run (consisting of multiple elements, including a spinning wheel), placed it opposite the box and equidistant from the participant, and said that the participant could also play with "this fun marble run," or could open the box and play with the marble run. In reality, the box was taped up with multiple layers of transparent tape. Thus, although this was not im- mediately obvious, the box was actually impossible to open. The experimenter turned her back toward the participant and walked into a different corner of the room, where she pretended to be absorbed in taking notes on her clipboard. She let the participant interact with the box and/or marble run for 2min before she returned. If the participant asked the experimenter for help or tried to interact with her in another way during this time, the experimenter stated in a neutral tone that she would be back in a bit, without responding to the help request.

Coding

We had two primary dependent variables: the difference score of children's confidence judgments (before [t1] vs after [t2] experiencing disagreement/agreement) and children's information search. To obtain each child's dif- ference score, we calculated the difference between that child's first and second confidence rating on the four-point scale (t2–t1). This resulted in a value between −3 (indicating a shift from "really sure" to "really not sure") and +3 (indicating a shift from "really not sure" to "really sure"). To obtain the information search

measure, we assessed how many seconds (out of the 2 minutes after the experimenter had turned around) children spent searching for additional information by engaging with the box that contained the blicket (note that per our pre- registration, searching for information was defined as either physically touching or visually attending to the box; however, there were no instances in either Experiment 1 or 2 in which children simply looked at the box without manually exploring it). Twenty-five percent of the information search data were coded for reliability by a second coder; interrater re- liability was excellent (ICC=0.99; Shrout & Fleiß, 1979; see Supplementary Material).

Data analysis

Children's difference scores and information search data were analyzed using linear regression models (via the stats package in R; R Core Team, 2021). Our two central predictions were that (1) children of all ages would reduce their overconfidence more in the disagreement relative to the agreement condition and that (2) children of all ages would search longer for information in the disagreement, compared to the agreement condition. We were also interested in potential developmental differences in the effects of disagreement on overconfidence and information search. Specifically, prior work has shown that between ages 4 and 6, children's overconfidence decreases (e.g., Hagá & Olson, 2017), and their explicit theory of mind reasoning abilities increase (e.g., Wellman & Liu, 2004). Given these prior findings, we wanted to be able to test whether, when confronted with disagreement, the older children in our sample would re- duce their overconfidence more and search longer for in- formation than the younger children in our sample. Thus, the main predictor in both models was the interaction between condition (disagreement vs agreement) and children's age (in years and months). Models also included children's gender and the order in which the hypotheses about blickets were mentioned (blue first and square second vs square first and blue second) as control predictors. To test for significance, we first compared each full model to a respective null model containing only the control predictors using likelihood-ratio tests (via the lrtest function from the lmtest package in R, Hothorn et al., 2015). When the full versus null model comparison was not significant, we included only the main effects for condition and age. To test the significance of individual predictors, we compared the full models with those of reduced models not containing these predictors using likelihood-ratio tests (via the drop1 function from the R stats package). Note that across experiments, none of our analyses revealed significant effects of the control predictors, so control predictor results are reported in the Supplementary Material. Additional analyses (e.g., those related to children's justifications) can also be found in the Supplementary Material.

Results

Children's confidence judgments

Initial belief and confidence at t1. Across conditions, almost all children immediately selected one of the two beliefs (square or blue). In line with prior work showing that children generally think shape is a good indicator of object category membership (e.g., Diesendruck & Bloom, 2003), children were somewhat more likely to think that square toys were blickets than that blue toys were blickets (disagreement condition: 60% “square”; agreement condition: 54%

"square"). Children's initial belief selection did not affect their initial confidence ratings or their change in confidence from t1 to t2 (see Supplementary Material).

On average, children in both conditions were fairly confident that their initial belief was correct, falling between "a little sure" and "really sure" (disagreement condition: M_{t1}=3.24 out of 4, SD_{t1} = 0.90; agreement condition: M_{t1} = 3.26 out of 4, SD_{t1} = 0.89; see Figure 6). These initial confidence ratings did not differ significantly across conditions ($t_{t1}(65.60) = .068$, p = .946). Given that the evidence that children had observed was completely ambiguous, this meant that children in both conditions exhibited overconfidence at t1. In line with prior research on children's overconfidence (e.g., Hagá & Olson, 2017a), we additionally found that older children were significantly less overconfident at t1 than younger children (see Supplementary Material for preregistered secondary hypothesis and results).

Changes in confidence from t1 to t2. Descriptively, from t1 to t2, children in the disagreement condition became less confident that their initial belief was correct (M_{t2} = 3.03, SD_{t2} = 0.95, $M_{difference\ score}$ = -0.21, $SD_{difference\ score}$ = 0.74), while children in the agreement condition became more confident that their initial belief was correct (M_{t2} = 3.40, SD_{t2} = 0.91; $M_{difference\ score}$ = 0.14, $SD_{difference\ score}$ = 0.65). T-tests revealed that these pre- versus post-differences were not significant (disagreement condition: $t_{t1\ vs\ t2}(34) = -1.30$, p = .201; agreement condition: $t_{t1\ vs\ t2}(32) = 1.65$, p = .109).

The preregistered linear model predicting children's difference scores from the interaction between condition and children's age was not significant when compared to its corresponding null model ($\chi^2(-3)=5.13$, p =.162), so we ran a model including only the main effects of condition and age, per our preregistration. This model revealed a significant effect of condition ($\beta = -.376$, 95% CI = [−.721, −.032], $\chi^2(1)=4.78$, p =.029): as expected, children expressed significantly less confidence in a belief formed based on ambiguous evidence following disagreement compared to agreement (see Figure 6). The effect of age was not significant ($\chi^2(1) = 0.002$, p =.964).

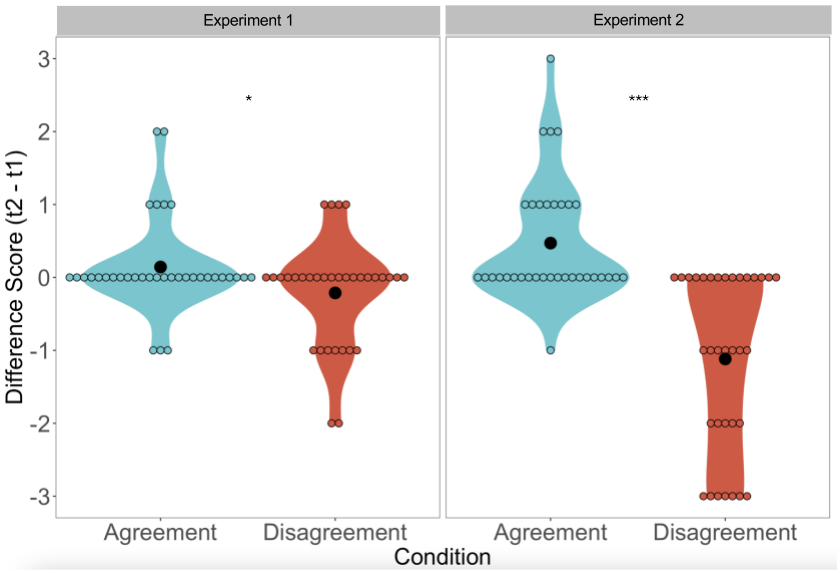
Children's information search

In the disagreement condition, 16% of children did not approach the box and played only with the marble run; the other 84% of children engaged in information search. Out of the children who searched, 69% searched for information and played with the marble run, while 31% engaged exclusively in information search. In the agreement condition, 26% of children played only with the marble run, while the other 74% of children searched for information. Out of these children, 81% both searched for information and played with the marble run, while 19% of children searched for information only. A two-sample test for equality of proportions revealed no significant difference between the proportions of children who searched (versus did not search) for information in the disagreement versus the agreement condition ($\chi^2(1) = 0.42$, p = .516).

On average, children in the disagreement condition spent 66.77 s searching (SD = 43.50 s), while children in the agreement condition searched for 48.17s (SD = 41.74 s). A t-test showed that this difference was not significant ($t(62.31)=1.77$, p = .082; see Figure 7). The preregistered linear regression model predicting children's information search from the interaction between condition and children's age was not significantly different from its corresponding null model ($\chi^2(-3) = 4.70$, p = .200), so we ran a model including only the main effects of condition and age. Neither condition ($\chi^2(1) = 2.02$, p = .15) nor age ($\chi^2(1) = 1.76$, p = .185) revealed significant effects.

Figure 6

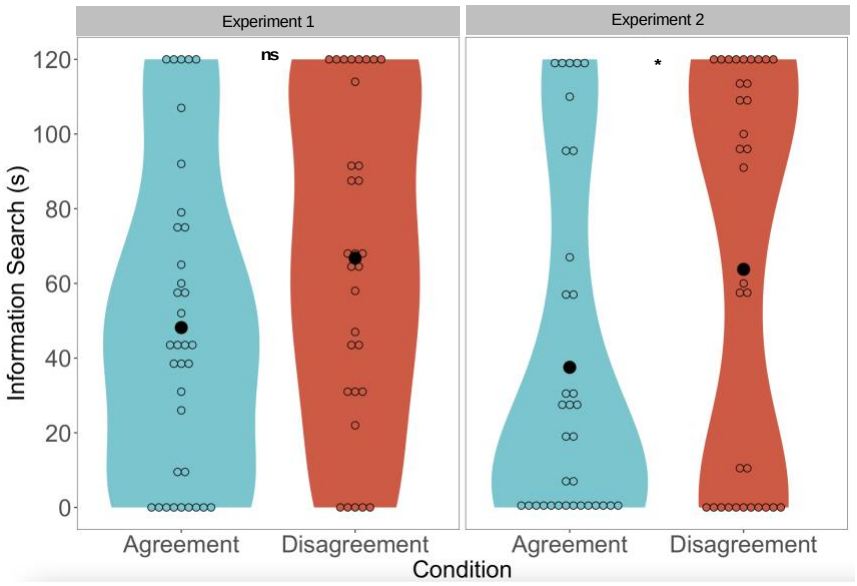
Children's difference scores in confidence (confidence at t2–confidence at t1).



Note. The figure shows children's change in confidence in the Disagreement and the Agreement conditions of Experiments 1 (left) and 2 (right). Solid dots are condition means; empty dots are individual data points.

Figure 7

Children's information search.



Note. Time children in the disagreement and the agreement condition spent searching for information in Experiments 1 (left) and 2 (right). Solid dots are condition means; empty dots are individual data points.

Discussion

Experiment 1 showed that young children who initially expressed unjustified confidence in a belief significantly reduced their overconfidence after experiencing disagreement (compared to agreement) with an adult. Importantly, although the youngest children in our sample were more overconfident than older children to begin with, disagreement led to similar confidence reductions across our tested age range. Children in the disagreement condition also searched longer for additional information than children in the agreement condition, but this difference was not significant.

Although children's confidence in the disagreement condition reduced significantly compared to the agreement condition, it was not significantly different from baseline ratings (t1), and remained in the "sure" range of the scale (i.e., above 3 on the four-point scale). Thus, children continued to exhibit overconfidence in their belief, even after disagreement. One possible explanation for why the disagreement did not have stronger effects is that it was not convincing enough: the confederate and the child had observed the same ambiguous evidence, so the confederate had no epistemic advantage over the child. In Experiment 2, we instead introduced the confederate as a "blicket expert," expecting that this would lead to a stronger reduction in overconfidence.

3.3 Experiment 2

Methods

Participants

In Experiment 2, we tested a new sample of 68 children; 34 in the disagreement condition and 34 in the agreement condition (M_{age} = 5.43 years, SD_{age} = 0.79, 33 girls, 35 boys). As in Experiment 1, children were from families with mixed SES backgrounds. Twenty-seven percent of children were white, 25% had multiple races or ethnicities, 15% were Hispanic, 13% were Asian, and 3% were African American (information on race or ethnicity was not provided by 17% of parents). Four additional children were excluded because the blicket machine malfunctioned. We had to exclude the information search data for one child (due to family interference), so the analysis of the information search data is based on data from the remaining 67 children. Children for Experiment 2 were tested between February and July 2022, and recruited, tested, and compensated as in Experiment 1.

Procedure

The procedure was identical to Experiment 1, except for one key change: In both conditions, after the experimenter had stated the goal of the game (i.e., to figure out which toys were blickets), she added, "I asked [confederate] to play with us today because (s)he knows a lot about blickets. (S)he knows more about blickets than anyone else I know. All of her/his friends call her/him a blicket expert because (s)he knows so much about blickets." The confederate confirmed this by saying "That's right I know a lot about blickets." This expertise manipulation was modeled after Koenig and Jaswal (2011). Afterward, the experiment continued exactly as in Experiment 1 (Figure 5).

Coding and data analysis

The data from Experiment 2 were coded and analyzed as in Experiment 1. Twenty-five percent of the information search data were coded by a second coder, and interrater reliability was excellent (ICC = 1; Shrout & Fleiss, 1979; see Supplementary Material). In addition, we conducted the following preregistered secondary analyses. First, in order to determine whether disagreement with an expert had stronger effects on children's overconfidence than disagreement with a confederate who had the same amount of limited knowledge as the child, we compared children's difference score data across experiments. The main predictor in this model was the interaction between experiment and condition, and age was included as an additional predictor. We were also interested to see how our two dependent variables were related. To test this, we filtered our dataset to include data from the disagreement condition only and predicted children's information search in the disagreement condition from the interaction between children's confidence at t2 and their age, including gender and counterbalance order as control predictors. We predicted that lower confidence ratings after experiencing disagreement with the confederate would predict longer information search.

Results

Children's confidence judgments

Initial belief and confidence at t1. In Experiment 2, children were about as likely to think that square toys were blickets as they were to think that blue toys were blickets (disagreement condition: 53% "square"; agreement condition: 53% "square"). As in Experiment 1, children's initial belief selection did not affect their confidence ratings, or their change in confidence ratings from t1 to t2 (see Supplemental Material).

Children in both conditions initially expressed strong confidence in their beliefs (i.e., they were overconfident; disagreement condition: M_{t1} = 3.53 out of 4, SD_{t1} = 0.79; agreement condition: M_{t1} = 3.09 out of 4, SD_{t1} = 1.08; see Figure 6). T1 confidence ratings did not differ significantly across conditions ($t(60.23) = -1.92$, $p = .060$). Again, we also found that the older children in our sample were initially less overconfident than the younger children (see Supplementary Material).

Changes in confidence from t1 to t2. Unlike in Experiment 1, the confidence of children in the disagreement condition significantly decreased from t1 to t2 (M_{t2} = 2.41, SD_{t2} = 1.26, $M_{difference\ score}$ = − 1.12, $SD_{difference\ score}$ = 1.20, 95% CI [−1.536, −0.699], $t(33)=-5.43$, p <.001, d =1.065); so much so, in fact, that children's average confidence ratings were no longer in the "sure" range of the scale. In contrast, confidence in the agreement condition significantly increased (M_{t2} = 3.56, SD_{t2} = 0.82, $M_{difference\ score}$ = .47, $SD_{difference\ score}$ = 0.83, 95% CI [0.183, 0.759], $t(33)=3.33$, p =.002, d =.489). The preregistered linear model predicting children's difference scores from the interaction between condition and age was significant compared to its null model ($\chi^2(-3) = 37.13$, p <.001). A test of the individual predictors revealed no significant effect of the interaction between condition and age ($\chi^2(1) = 3.32$, p =.068). A model containing only the main effects for condition and age revealed a highly significant effect of condition, as expected ($\beta = -1.590$, 95% CI [−2.086, −1.094], $\chi^2(1) = 33.08$, $p < .001$). The effect of age was not significant ($\chi^2(1) = 0.86$, $p = .354$).

When comparing the difference score data across experiments, the full model predicting children's difference scores from the interaction between experiment and condition was significant compared to its null model ($\chi^2(-4) = 53.18$, p <.001). Specifically, there was a

significant effect of the experiment * condition interaction (β = −1.256, 95% CI [−1.855, −.656], $\chi^2(1) = 13.28$, $p < .001$). As shown in Figure 6, this effect was due to the condition difference being larger in Experiment 2, where children disagreed with an expert, than in Experiment 1, where children disagreed with a confederate who had the same amount of limited knowledge as the child. The effect of age was not significant ($\chi^2(1) = 0.51$, $p = .475$).

Children's information search

In the disagreement condition, 33% of children did not approach the box containing the blicket; the other 67% of children did. Out of the children who searched, 59% searched for information and played with the marble run, while 41% engaged exclusively in information search. In the agreement condition, 38% of children did not search, while 62% of children did. Out of these children, 62% searched for information and played with the marble run, while 38% of children searched for information only. A two-sample test for equality of proportions revealed no significant difference between the proportions of children who searched (versus did not search) for information in the disagreement versus the agreement condition ($\chi^2(1) = 0.38$, $p = .536$).

On average, children searched for 63.76s in the disagreement condition (SD = 53.66s) and for 37.53s in the agreement condition (SD = 46.02s; see Figure 7). A t-test showed that this difference was significant ($t(62.91) = 2.14$; p = .036). The preregistered model including the interaction between condition and age was not significantly different from its null model ($\chi^2(-3)=5.00$, $p =.172$), but a model including only the main effects for condition and age revealed a significant effect of condition in line with our prediction (β = 26.613, 95% CI [2.263, 50.964], $\chi^2(1) = 4.78$, $p = .029$). The effect of age was not significant ($\chi^2(1) = 0.18$, $p = .668$).

Relationship between children's confidence judgments and information search

The model predicting the information search of children in the disagreement condition from the interaction between children's confidence at t2 and their age was significant compared to its null model ($\chi^2(-3) = 9.86$, $p = .020$). However, a test of the individual predictors revealed no significant effect of the interaction ($\chi^2(1) = 0.08$, $p = .774$) or the control predictors (see Supplementary Material). A model including only the main effects of children's confidence at t2 and their age showed a significant effect of children's t2 confidence (β = 23.381, 95% CI [8.639, 38.123], $\chi^2(1) = 9.78$, $p = .002$), revealing that children with lower confidence ratings at t2 searched longer for information. Age was not significant ($\chi^2(1)=1.71$, $p =.191$).

3.3 General Discussion

Across two preregistered studies, we tested whether experiencing disagreement – as opposed to agreement – would (1) reduce 4- to 6-year-olds' overconfidence and (2) motivate them to search for additional information regarding the correct answer. Our findings – particularly those of Experiment 2 – confirmed these predictions, providing the first evidence that disagreement can reduce overconfidence in young children and prompt their search for information in adaptive ways that could benefit cognitive development.

Although younger children in our sample were initially more overconfident than older children, experiencing disagreement reduced children's overconfidence similarly across ages. Thus, unlike other interventions, such as feedback (e.g., Buehler et al., 2023), disagreement

seems to be an effective tool in helping young children better calibrate their certainty. Importantly, while disagreement with an adult confederate who had the same amount of limited prior knowledge as the child (in Experiment 1) reduced children's overconfidence to some extent, only disagreement with an adult confederate who was introduced as an expert (in Experiment 2) reduced children's overconfidence significantly compared to baseline (and significantly more so than in Experiment 1). This suggests that children were sensitive to the epistemic status of whom they were disagreeing with (see Koenig & Jaswal, 2011) and did not simply reduce their overconfidence in response to an adult's greater dominance or social status. In future work, it would be interesting to examine whether children reduce their overconfidence similarly when disagreeing with naïve versus experienced peers of the same age.

We found not only that disagreement can lead children to scale back from their overconfidence but also that it can have tangible consequences, in leading children to search for additional information related to the question at hand (descriptively in Experiment 1 and statistically in Experiment 2). In showing that a central element of children's social worlds – disagreement – can prompt children's exploration in adaptive ways, our findings extend prior work on social-communicative cues that can influence children's exploration, such as pedagogical instruction (Bonawitz et al., 2011), and observing other people's surprising actions (Stahl & Woods, 2022) or emotional expressions (Wu & Gweon, 2021).

Together, our findings bridge research on explicit and implicit metacognition (see Goupil & Kouider, 2019). Prior research has usually either focused on explicit metacognition and revealed that young children are overconfident when asked to make explicit verbal confidence judgments (e.g., Hagá & Olson, 2017a); or it has focused on implicit metacognition and revealed that even 2-year-olds demonstrate metacognitive sensitivity in implicit behavioral paradigms (e.g., Hembacher et al., 2020), for example, by searching longer for information in situations of greater uncertainty. The setup of the current experiments allowed us to measure explicit and implicit forms of metacognition within the same paradigm. We found in Experiment 2 that across our tested age range, children's stated confidence in their belief predicted their search behavior. Our findings thus raise the possibility that children's explicit confidence judgments can influence their intuitive and adaptive behavioral responses to uncertainty (e.g., with children searching less long for information after stating that they were “really sure”).

Together, our results not only add to the burgeoning literature showing positive impacts of disagreement on children's reasoning (e.g., Li & Tomasello, 2022) but also suggest that reduced overconfidence may be an important psychological mechanism underlying some of these effects. For example, reduction in confidence may prompt the rational belief revision observed in prior studies (e.g., Langenhoff et al., 2023; Schleihauf et al., 2022) and generate information search, as in the current experiments. But why might disagreement produce these effects (both in our experiments, and more generally)? We have suggested that disagreement reduces children's overconfidence and prompts their exploration because it highlights the presence of alternative perspectives, thereby alerting children to the possibility that they might be mistaken. However, it is also possible that children might make different inferences about the personality of another person who disagrees versus agrees with them. In our experiments, for instance, children might have construed the disagreeing confederate as being unfriendly, which, in turn, may have emotionally impacted children and reduced their confidence. Although we did not directly measure children's emotional responses or their social evaluation of the confederate, children arguably would have been more likely to perceive the confederate as unfriendly in

Experiment 1, where the confederate lacked any additional prior knowledge to support their contrasting belief, compared to in Experiment 2, where the confederate had a justification for their differing belief due to being more knowledgeable. Yet, we observed more pronounced effects in Experiment 2, suggesting that it was not the emotional impact of experiencing disagreement that influenced children's responses.

A related question to consider is what motivates children's increased information search after experiencing disagreement. We have suggested that children who have experienced disagreement search longer because they are less certain and thus more curious about the correct answer. This idea is supported by our finding that in Experiment 2, lower confidence at t2 predicted children's information search in the disagreement condition. However, given that we observed significant variability in children's information search in both Experiments 1 and 2 (see Figure 7), it is likely that other factors – in addition to increased curiosity resulting from the disagreement – influenced children's information search behaviors. For instance, in both studies, some children said they were "really sure" at t2, and nonetheless engaged in information search for the entire two-minute duration. These children's search may have been driven not by reduced confidence in their initial belief but by a desire to validate or affirm their existing belief, and to prove that they were right. Individual differences in persistence represent yet another factor that may have affected the extent of children's information search. As mentioned, opening the box to determine the true answer was very difficult (and in fact impossible), and we know from prior work that there is significant individual variation in how long young children persist at difficult tasks (Banerjee & Tamis-LeMonda, 2007; Martin et al., 2013; Mokrova et al., 2013; see also Leonard et al., 2021). In future work, the influence of these different factors could be determined by including a baseline measure of children's information search.

Including a baseline measure of children's confidence could also help to identify cases in which a decrease in confidence does not lead to an increase in information search. In our experiments, we illustrated how a reduction in overconfidence can stimulate greater information search by evoking curiosity about the correct answer. However, we expect that there will be situations in which confidence diminishes without leading to increased exploration. One example, as mentioned above, is when an individual is highly motivated to explore, not due to reduced confidence but rather because of their high confidence and their motivation to prove to others that they are right. Conversely, an individual might exhibit such low confidence that their inclination to explore becomes suppressed, as they may not believe that their own knowledge or information-seeking efforts will yield new insights (i.e., due to reduced self-efficacy). A fruitful direction of future research will be to consider how confidence and information search can be associated or dissociated across different situations.

Our studies provide initial empirical support for social accounts of reasoning which propose that many higher human reasoning capacities might be "at their best" in social contexts (Dutilh Novaes, 2018; Mercier & Sperber, 2011; Köymen & Tomasello, 2020; O'Madagain & Tomasello, 2021). Notably, this proposal has been extended toward explaining the development of metacognitive reasoning, in particular (Heyes et al., 2020), which includes the ability to monitor one's (un)certainty and act upon it in reasonable ways. While the idea that individual reasoning is facilitated within social contexts – and specifically, within contexts of disagreement – is intriguing, it requires further study. Most importantly, studies are needed to probe whether the experience of disagreement has unique effects on reasoning, compared to being exposed to alternative hypotheses or additional information in a non-social manner (e.g., by introducing

conflicting vs consistent beliefs from a non-social source, such as a robot or a computer). We are aware of two prior studies (Doise et al., 1975; O'Madagain et al., 2022) that have compared children's reasoning in a disagreement condition with their reasoning in a non-social comparison condition: both found that children benefited more from social disagreement. However, additional studies are needed to understand if and why disagreement may have these unique effects.

Our findings could hold significant potential for interventions aimed at fostering intellectual humility (Leary et al., 2017) and promoting learning (Baer & Kidd, 2022). To determine the generalizability of our findings and establish the power of disagreement-based interventions, it will be important to understand how the effects of disagreement observed here might transfer to other contexts and domains. Although our focus in the present studies was on young children, who are arguably particularly good targets of intervention (due to their overconfidence and egocentrism), disagreement-based interventions could be valuable across the lifespan, given that overestimating one's own knowledge and under-appreciating others' perspectives constitute some of the most problematic human tendencies in contemporary societies (Zmigrod et al., 2019).

Importantly, although our findings illustrate the positive consequences of disagreement, disagreement can clearly also lead to negative consequences. Disagreements are inherently social processes, and so an individual's social motives, biases, and stereotypes can likely influence how they respond to a disagreement. For example, when a person experiences disagreement regarding a belief that really matters to them, they will likely be more hesitant to reduce their confidence in that belief (see Kahan, 2017; Oktar & Lombrozo, 2022). In fact, some evidence suggests that such disagreements can sometimes even strengthen individuals' prior beliefs (e.g., Nyhan & Reifler, 2010). Moreover, constructive disagreements require that the disagreeing parties treat each other with respect and give each other the credibility they deserve. When individuals are not receptive to the alternative beliefs of another due to prejudices or stereotypes, disagreement will likely fail to elicit positive consequences (see Fricker, 2007). Identifying the features of disagreement that support intellectual humility and learning will be crucial for the effective design of interventions that facilitate learning and mutual understanding.

Chapter 4: Disagreement encourages children's rational belief revision

4.1 Introduction

From Gram Sabhas, one of the world's largest deliberative institutions in India, to the Kgotla, a traditional community meeting in rural Botswana, people around the world reason together to make collective decisions (Sen, 2003). Public deliberation can be a powerful tool to overcome polarization, counteract populism, and make better judgments – under the right conditions. On the individual level, an important requirement for public deliberation is that people respond to disagreement in reasonable and intellectually humble ways. Sound public decisions can only be reached if people consider points of view that differ from their own, and respond rationally when confronted with differing perspectives (Habermas, 1997).

But what constitutes a rational response to disagreement? The answer depends on what the disagreement is about. When disagreeing about preferences or values (for example, about the best restaurant in town), the ideal thing to do might be to find a compromise or to "agree to disagree" (see Amemiya et al., 2021; Fisher et al., 2017; Foushee & Srinivasan, 2017; Sen, 2003). But people also disagree about issues that have a single, objectively correct answer (such as the number of restaurants in town). In these situations, a reasonable individual should consider which of the two diverging beliefs is supported by stronger evidence. If the individual has good reason to think that the evidence supporting their belief is stronger than that of the person they are disagreeing with, they should maintain their initial belief; if the other person's evidence is stronger, they should adopt the other person's view. Sometimes, however, the evidence supporting the individual's own belief versus the disagreeing other person's belief will be equally strong. By many epistemological accounts, an individual who finds themselves in this kind of disagreement should suspend judgment, and acquire additional evidence before coming to a conclusion (Frances, 2014; Friedman, 2017). Importantly, as mentioned in Chapter 1, suspension of judgment is not merely the absence of an attitude; instead, it's widely regarded by epistemologists as an explicit metacognitive stance, in which individuals introspect on their own beliefs, acknowledging that the evidence falls short of confirming it, and consciously recognize their uncertainty about the truth (Crawford, 2022; Friedman, 2013, 2017; Turri, 2012).

Here, we investigated children's developing ability to respond to disagreement in the above-described three distinct ways across two studies. Research illuminating the developmental origins of responses to disagreement has both theoretical and practical implications. On a theoretical level, some have argued that the human ability to reason is best conceptualized not as an individual skill, but as a fundamentally social skill, thus defining reasoning as the social practice of giving and asking for reasons in interpersonal discourse (Heyes, 2012; Köymen & Tomasello, 2020; Mercier & Sperber, 2011; O'Madagain, 2019; O'Madagain & Tomasello, 2021; Schleihauf et al., 2022; Tomasello, 2020). By these accounts, the ability to adjust one's belief appropriately in light of disagreement constitutes a critical hallmark of a rational reasoner. On a practical level, such research can inform interventions aimed at fostering argumentation skills and intellectual humility (see Danovitch et al., 2019; Hagá & Olson, 2017a; Porter & Schumann, 2018). Fostering these skills, in turn, can contribute to facilitating a healthy public discourse.

Much prior work in Developmental Psychology has looked at children's developing ability to understand and evaluate disagreements among third parties (e.g., Amemiya et al., 2021; Birch et al., 2008; Chen et al., 2013; Corriveau et al., 2009; Foushee & Srinivasan, 2017;

Heiphetz & Young, 2017; Koenig, 2012; Mercier et al., 2018; Scofield & Behrend, 2008). One central finding from this literature is that, when presented with a disagreement between other agents, even preschoolers are sensitive toward which of two opposing beliefs is supported by stronger evidence (quantitatively or qualitatively), and preferentially adopt that belief. For example, 3- to 4-year-olds preferentially accept a claim that is supported by a unanimous majority over one that is supported by a lone dissenter (e.g., Chen et al., 2013; Corriveau et al., 2009). Around age 4, children also prefer to adopt the beliefs of those who support their claims with strong reasons over the beliefs of those who provide only weak reasons (e.g., Koenig, 2012; Mercier et al., 2018). Thus, when evaluating disagreements between third parties, children respond reasonably from early on.

However, for understanding the development of children's intellectual humility and their ability to become rational contributors to public discourse, it is particularly important how children themselves respond to disagreement. This is what we study here. Prior studies investigating children's responses to first-person disagreements have focused on whether children maintain their initial belief or adopt the conflicting testimonial evidence from another person (Hagá & Olson, 2017b; Jaswal, 2010; Jaswal et al., 2014; Ma & Ganea, 2010; Miosga et al., 2020; Schleihauf et al., 2022; Young et al., 2012). Interestingly, these studies show that 2- to 3-year-olds sometimes adopt the belief of a disagreeing other, even if that belief conflicts with an event the child has just witnessed (Jaswal, 2010; Jaswal et al., 2014; Ma & Ganea, 2010). Thus, young children appear to often use disagreement as an indicator that they are mistaken, and employ a "when somebody disagrees with you, adopt their view" – strategy (see also Hagá & Olson, 2017b).

Slightly older children, in contrast, appear to consider both the quality of the evidence for their initial belief, as well as the quality of the evidence for the belief of a disagreeing other when deciding whether or not to change their minds after a disagreement—at least in situations where it is easy to see which of the two opposing belief is supported by stronger evidence (Miosga et al., 2020; Young et al., 2012; Schleihauf et al., 2022; see also Bridgers et al., 2016). Finally, research with even older children suggests that, by age 7 – sometimes referred to as the "age of reason" (see Tomasello, 2020) – children reliably compare the strength of the evidence supporting their own claim versus the claim of a disagreeing other, even in more complex contexts. For example, 7-year-olds selectively maintain their initial belief or adopt another person's belief, depending on which belief is supported by first- versus second-hand evidence (Köymen & Engelmann, 2022; Köymen & Tomasello, 2018; see also Morgan et al., 2015).

Taken together, prior research suggests that there appear to be considerable improvements with age in children's ability to respond to first-person disagreements. However, this work leaves two key questions unanswered. First, prior studies have exclusively focused on children's ability to selectively maintain their initial belief or adopt the belief of a disagreeing other (e.g., Jaswal, 2010; Jaswal et al., 2014; Ma & Ganea, 2010; Schleihauf et al., 2022). Yet, as explained above, an additional hallmark of rational reasoning is to suspend judgment when the evidence supporting one's own belief and a disagreeing other's belief are equally strong (Frances, 2014; Friedman, 2017). The first goal of our studies was thus to investigate whether, when disagreeing with another person, children would (1) maintain their initial belief when the evidence supporting their own belief was stronger, (2) adopt the other's belief if their own evidence was weaker, and (3) suspend judgment when the evidence supporting the two conflicting beliefs was equally strong. Second, prior work focusing on children's responses to

first-person disagreement has rarely compared children of different age groups within the same paradigm. Thus, it is unclear whether the observed age-related increases in reasonable responses to disagreement reflect genuine develop- mental change. The second goal of our studies was thus to investigate potential developmental differences in children's responses to disagreement. Specifically, we compared how children of two different age groups responded to disagreement: 4- to 6-year-olds and 7- to 9-year-olds (we did not test children below age 4 be- cause piloting with 3-year-olds had revealed that our task was too difficult for these children).

To answer the two questions described above, we presented children and adults (for comparison) with a story in which a character's pet had run away, and asked participants to find out where the pet was hiding. Participants first acquired evidence by asking informants, and were then asked to state their belief about where the pet went. Next, participants learned that another agent – who had also consulted informants on the whereabouts of the pet – disagreed with the participant. Finally, we assessed participants' belief for a second time. We were interested in whether participants would maintain their initial belief, suspend judgment by saying they were "not sure" and "needed more information," or adopt the other agent's belief. Across conditions, we varied whether the evidence supporting the participants' belief was stronger than, weaker than, or equal to the evidence supporting the belief of the disagreeing agent.

Our main indicator of reasonableness was whether participants would maintain their initial belief in the stronger evidence condition; adopt the other agent's belief in the weaker evidence condition; and suspend judgment in the equal evidence condition more often than in the respective other two conditions. We chose to compare participants' responses in this way because we assumed that there would be baseline differences in how attractive the different response options would be, and that these base- line differences might vary between age groups. Since our main focus was on whether children's and adults' responses vary reasonably as a function of the different epistemic conditions despite these underlying baseline differences, our preregistered main analyses compared participants' responses across conditions rather than within conditions.

Based on prior work (e.g., Köymen & Tomasello, 2018), we expected that by age 7, children would respond to disagreement in these ways, similarly to adults. In contrast, we expected that younger children would diverge from older children and adults in two ways. First, although previous research shows that children begin to compare the quality of the evidence supporting their own belief versus the belief of a disagreeing other when deciding whether to change their minds after a disagreement (Miosga et al., 2020; Schleihauf et al., 2022; Young et al., 2012), prior work also suggests that they are generally more open to the suggestions of others than older children and adults (see Bridgers et al., 2016; Hagá & Olson, 2017b). We thus expected that, across conditions, 4- to 6-year-olds would be more likely to adopt the belief of the person they were disagreeing with. Second, prior research suggests that the ability to suspend judgment might be difficult for young children, particularly if it requires them to make a metacognitive judgment about their own uncertainty (Butterfield et al., 1988; Coughlin et al., 2015; Hagá & Olson, 2017b; Lipko et al., 2009; Lipowski et al., 2013; Perner, 2012). In Experiment 1, suspension of judgment was assessed with an explicit meta- cognitive judgment; participants suspended judgment by indicating that they were "not sure" and needed "more information." Accordingly, we predicted that 4- to 6-year-olds in Experiment 1 would be less likely to sus- pend judgment than the other age groups. Experiment 2 explored whether a more implicit operationalization of suspension of judgment would make it easier for 4- to 6-year-olds to respond rationally to the ambiguous evidence presented to them in the equal evidence con-

dition (Hembacher et al., 2020; Lapidow et al., 2022; Lyons & Ghetti, 2012). Hypotheses, exclusion criteria, analyses, and design for both experiments were preregistered (Experiment 1: https://aspredicted.org/cw9wa. pdf; Experiment 2: https://aspredicted.org/cw9wa.pdf).

4.2 Experiment 1

Methods

Participants

A total of 114 participants completed Experiment 1: 38 younger children aged 4.0–6.11 (16 4-year-olds, eight 5-year-olds, 14 6-year-olds, M_{age} = 5.46 years, SD_{age} = 0.97 years, 50% female), 38 older children aged 7.0 – 9.11 (17 7-year-olds, 20 8-year-olds, one 9-year-old, M_{age} = 8.10, SD_{age} = 0.55, 50% female) and 38 adults (M_{age} = 31.45, SD_{age} = 13, 50% male, 45% female, 5% agender or non-binary). Children were recruited from a database of families who had signed up to participate in child development studies. Children were predominantly white (45%), followed by Asian (8%), Latinx and white (8%), and Latinx (7%). Sixteen percent of families did not indicate their child's race or ethnicity. Upon study completion, children received a certificate. Two additional children participated but were excluded in accordance with our exclusion criteria because they stopped responding after the second trial. Data for children were collected between June and August 2020.

Adults were recruited via Prolific. Participation was restricted to participants based in the United States with a prior approval rate higher than 90%. Adults were mostly white (66%), followed by African or African American (13%), Asian (11%), Latinx (8%), or white and Latinx (2%). Adults were compensated with $1.60; their data were collected in December 2020.

Our total sample size was determined based on an a priori power analysis. We calculated an estimate of the sample size required to find a small to medium effect (f = 0.2) of the interaction between our predictors (condition and age group) on participants' responses to the second belief assessment with 90% power. This analysis resulted in an optimal sample size of N = 111.

Design

The experiment had three within-subjects conditions – corresponding to three different test trials – which differed in the relative strength of evidence that participants were exposed to before encountering a disagreement with another agent. In the stronger evidence condition, participants received testimony from three informants who told participants where the pet went. The other agent received testimony from one informant. In the weaker evidence condition, participants received testimony from one informant and the other agent received testimony from three informants. In the equal evidence condition, both the participant and the other agent received testimony from two informants. The order in which participants were presented with the conditions was randomized.

Procedure

Before the experiment, children were asked if they wanted to participate, and only participated if both they and their parents agreed. Adults provided written consent. Testing took place online; experiments were implemented in Qualtrics. Test sessions with children were

conducted by an experimenter via Zoom; adults participated self-directed, via Prolific. An overview of the experimental procedure is depicted in Figure 8.

Exposure to evidence. During the experiment, participants either listened to an experimenter reading (children) or themselves read (adults) a picture book about Emma and her bunny. Participants took part in one practice trial and three test trials. Each trial began as follows: Emma's bunny ran away and Emma was sad. One of Emma's friends offered to find the bunny and encouraged the participant to help them, suggesting that the friend and the participant could each ask some people if they saw where the bunny went. There were always two options of where the bunny could have gone, for example toward the house or toward the bridge (presentation of stimuli was counter- balanced). Next, participants were exposed to one, two, or three informants (number varied based on condition), who, one after another, told the participant where the bunny had gone. Informants always provided a consistent response, for example, all informants said the bunny went toward the house. In order to avoid carryover effects between trials, participants "interacted" with each character only once. That is, participants were exposed to a different friend of Emma's and to different informants in each trial.

Initial belief assessment. After being exposed to the statements of the inform- ants, the participant was prompted to state their belief about where the bunny went for a first time (e.g., in the weaker evidence condition, "So one person told you the bunny went to the house. What do you think? Do you think the bunny went to the house or do you think the bunny went to the bridge?"). Children received feedback on this initial belief assessment. If a child responded differently than what the informant(s) had told them, the experimenter repeated the test question. If a child responded incorrectly for a second time, the experimenter corrected the child (e.g., "remember, one person told you the bunny went toward the house. So I think the bunny went toward the house."). If a participant responded incorrectly to the first belief assessment (after feedback for children), the rest of the trial was skipped and the participant was directed to the subsequent trial. As specified in our preregistration, these trials were excluded from analyses. Moreover, if a participant responded to the initial belief incorrectly on more than one test trial, their data were excluded completely.

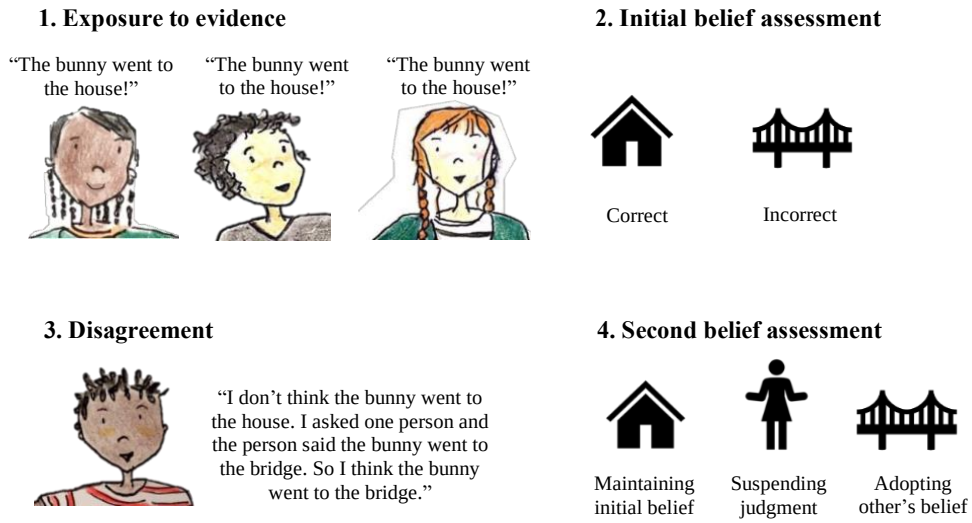
Disagreement and second belief assessment. The first trial was a practice trial that familiarized participants with the structure of the task. In the practice trial, the belief of Emma's friend confirmed the participant's initial belief. In contrast, in the three subsequent test trials, Emma's friend disagreed with the participant, and justified their belief by referring to how many in- formants they asked. For example, in the weaker evidence condition, Emma's friend said "I don't think the bunny went toward the house. I asked three people and they said the bunny went toward the bridge. So I think the bunny went toward the bridge."

After this disagreement occurred, the participant's belief about where the bunny went was assessed for a second time. For example, in the weaker evidence condition, the experimenter said "one person told you the bunny went toward the house. And you said the bunny went toward the house. And three people told [friend's name] the bunny went toward the bridge. And [friend's name] said the bunny went toward the bridge. What do you think now? Where do you think the bunny went?". The response options in this case were "toward the house", which represented maintaining one's initial belief, "toward the bridge", which represented adopting the friend's belief, and "I am not sure. I need more information.", which represented suspension of judgment. The response options were illustrated with pictures, for example, with a picture of the house and

a picture of the bridge. The “suspension of judgment” response option was illustrated with an icon of a person shrugging their shoulders in an “I don't know” gesture.

Figure 8

Overview of the experimental procedure for Experiment 1. The condition depicted here is the stronger evidence condition.



Data analysis

We used logistic linear mixed regression models (lme4 package in R, Bates et al., 2012) with a binomial distribution (logit link) to analyze participants' responses to the second belief assessment. In our primary preregistered analyses, the main predictor was the interaction between condition (stronger vs. weaker vs. equal evidence condition) and age group (adults vs. older children vs. younger children). In addition, in response to a comment from a reviewer, we also investigated the effects of age as a continuous predictor on children's responses. In these models, the main predictor was the interaction between condition and age (in years and months). If possible, all models included gender (female vs. male vs. nonbinary for adults) and trial number (1–3, z-transformed, to avoid convergence issues) as control predictors, random slopes for condition (mean- centered) and trial number (z-transformed), as well as a random intercept for subject. If a model did not converge, we removed correlations between random effects, then removed control predictors (first gender and then trial number), and finally removed random slopes (first trial number and then condition).

For our primary preregistered analyses, as well as for the analyses including age as a continuous predictor, we ran three different types of regression models with binary outcome measures: in the first type of model, we predicted participants' propensity to maintain their initial belief. For these models, responses were recoded as “maintaining one's initial belief” versus “other” (which in this case included suspension of judgment and adopting the friend's belief). In the second type of model, we predicted participants' propensity to adopt the friend's belief (recoding responses to “adopting the friend's belief” vs. “other”), and in the third type of model, we predicted participants' propensity to suspend judgment (recoding responses to “suspension of judgment” vs. “other”). For our primary preregistered analyses, we included data from the whole

sample first (with the interaction between condition and age group as our main predictor), and then looked at the effect of condition on each age group in a second step. For our analyses including age as a continuous predictor, we included all data from children.

To establish the significance of the full models (Schielzeth & Forstmeier, 2009), we compared the deviance of the full models with those of the null models containing only the control predictors, the intercept, and the random slopes, using likelihood-ratio tests (Dobson & Barnett, 2018). To test the significance of the interactions between condition and age, we compared the deviance of the full models with those of corresponding reduced models not containing the interactions, using likelihood-ratio tests. We used the R package emmeans (Lenth et al., 2018) to conduct post hoc analyses (Tukey method), when necessary. All materials for Experiments 1 and 2, including data, analysis scripts, and printouts of the Qualtrics surveys, are available here: https://osf.io/z2tbx/?view_only=de622 9ef7a2e4e37a1b9fc9b3e540b9c.

Results

We organize this section according to our three pre- registered predictions about how a reasonable person should respond to disagreement. We analyze whether (1) participants maintained their initial belief more often in the stronger evidence condition, (2) adopted the belief of the disagreeing other more often in the weaker evidence condition, and (3) suspended judgment more often in the equal evidence condition than in the respective other two conditions. As specified in our preregistered exclusion criteria, our analyses were based on trials in which participants responded correctly to the initial belief assessment (after feedback for children). In total, participants failed to respond correctly to the initial belief assessment on only five trials (adults: three trials in the weaker evidence condition; children: one trial in the stronger evidence condition, one trial in the weaker evidence condition). An overview of how participants responded in the different conditions in Experiment 1 and Experiment 2 is provided in Table 1.

Table 1

Frequencies with which participants in Experiment 1 and 2 responded to the disagreement by maintaining their initial belief, adopting the other's belief, or suspending judgment; separately for age group and condition.

Age group	Response	Experiment 1						Experiment 2					
		Stronger evidence		Weaker evidence		Equal evidence		Stronger evidence		Weaker evidence		Equal evidence	
		f	SE	f	SE	f	SE	f	SE	f	SE	f	SE
4-6 years	Maintain	.43	.08	.08	.04	.29	.07	.4	.07	.22	.06	.24	.06
	Adopt	.41	.08	.79	.07	.50	.08	.36	.07	.59	.07	.35	.06
	Suspend	.16	.06	.13	.05	.21	.07	.25	.06	.2	.06	.41	.07
7-9 years	Maintain	.76	.07	.03	.02	.29	.07						
	Adopt	.13	.05	.76	.07	.18	.06						
	Suspend	.11	.05	.22	.07	.53	.08						
Adults	Maintain	.84	.06	0	0	.08	.04	.82	.05	0	0	.1	.04
	Adopt	.05	.04	.83	.06	.13	.05	.04	.03	.84	.05	.1	.04
	Suspend	.11	.05	.17	.06	.79	.07	.14	.05	.16	.05	.8	.06

Note. f indicates relative frequencies, SE indicates the 95% Standard Error.

Did participants maintain their initial belief when their own evidence was stronger?

For analyses involving age group as a categorical variable, we could not fit a model to our whole dataset because there was not enough variance in the adult data: in the stronger evidence condition, adults maintained their initial belief 84% of the time, but they never did so in the other two conditions (see Table 1). A model including the data from children was highly significant when compared with the null model ($\chi^2(7) = 83.05$, $p < .001$). Specifically, we found a significant interaction between condition and age group ($\chi^2(2) = 8.02$, $p = .018$), while gender ($\chi^2(1) = 1.03$, $p = .310$) and trial number ($\chi^2(1) = 0.84$, $p = .360$) had no effect. To further investigate the interaction between condition and age group, we looked at the effects of condition on younger and older children in separate regression models. In both models, condition was a significant predictor (older children $\chi^2(2) = 49.43$, $p < .001$; younger children $\chi^2(2) = 25.99$, $p < .001$). Post hoc pairwise comparisons revealed that both age groups were more likely to maintain their initial belief in the stronger evidence condition than in the equal evidence condition (older children $p = .009$; younger children $p < .001$), and in the stronger evidence condition compared to the weaker evidence condition (older children $p < .001$; younger children $p = .004$). Thus, as groups, both younger (i.e., 4- to 6-year-olds) and older (i.e., 7- to 9-year-olds) children responded "reasonably" in the sense that they maintained their initial belief more often when their own evidence was stronger than that of Emma's friend, compared to when their own evidence was weaker or equally strong. Note that for the above-mentioned, as well as all additional analyses reported throughout the paper, we found the same trends when including data from participants first trials only. Details on first trial-analyses for Experiments 1 and 2 can be found in the supplementary materials. Further supporting this general pattern, maintaining one's initial belief was the most common response in all age groups when comparing responses within the stronger evidence condition. However, while the difference between maintaining one's initial belief and the two other response options was pronounced among older children and adults, 4- to 6-year-olds were almost as likely to adopt the other character's belief (see Table 1).

When analyzing the data from children including age as a continuous predictor, we had to take out the control predictors and the random slope for the model to converge. The full model was significant compared to the null model, ($\chi^2(5) = 35.57$, $p <.001$), revealing a significant interaction between condition and age ($\chi^2(2) = 18.26$, $p <.001$). As Figure 9 shows, the condition difference first emerged in 5-year-olds, and was increasingly clear-cut in older children. It appears that 4-year-olds did not maintain their belief more often in the stronger evidence condition than in the other two conditions.

Did participants adopt the other's belief when their own evidence was weaker?

When analyzing the data including age group as a cat- egorical predictor, the full model was significant when compared with the null model ($\chi^2(8) = 140.35$, $p < .001$). Specifically, the age group * condition interaction was significant ($\chi^2(4) = 12.41$, $p = .015$). Gender ($\chi^2(3) = 6.03$, $p = .110$) and trial number ($\chi^2(1) = 1.48$, $p = .224$) did not have a significant effect on participants' responses. We followed up on this interaction with separate regression models, and found that condition was a significant predictor in all three age groups (adults $\chi^2(2) = 100.72$, $p < .001$; older children $\chi^2(2) = 78.12$, $p < .001$, younger children $\chi^2(2) = 26.96$, $p < .001$). Specifically, all age groups adopted the friend's belief significantly more often in the weaker evidence condition than in the stronger evidence condition (adults $p < .001$; older children $p < .001$; younger children $p < .001$), and in the weaker evidence condition compared to the equal

evidence condition (adults $p < .001$; older children $p = .032$; younger children $p = .003$). Thus, as predicted, participants adopted the belief of Emma's friend more often when their evidence was weaker than that of Emma's friend, compared to when their own evidence was stronger or equally strong. Indeed, within the weaker evidence condition, all three age groups were more likely to adopt the other character's belief than to maintain their initial belief or to suspend judgment.

When analyzing children's responses including age as a continuous predictor, we had to take out the control predictor for the model to converge. Compared to the null model, the full model was significant ($\chi^2(5) = 74.06$, $p < .001$). Specifically, there was a significant interaction between condition and age ($\chi^2(2) = 10.27$, $p = .006$). As Figure 9 shows, children of all ages adopted the other's belief more often in the stronger evidence condition than in the other conditions, but became slightly more likely to do so with increasing age. In addition, older children also became less likely to adopt the other's belief in the other two conditions.

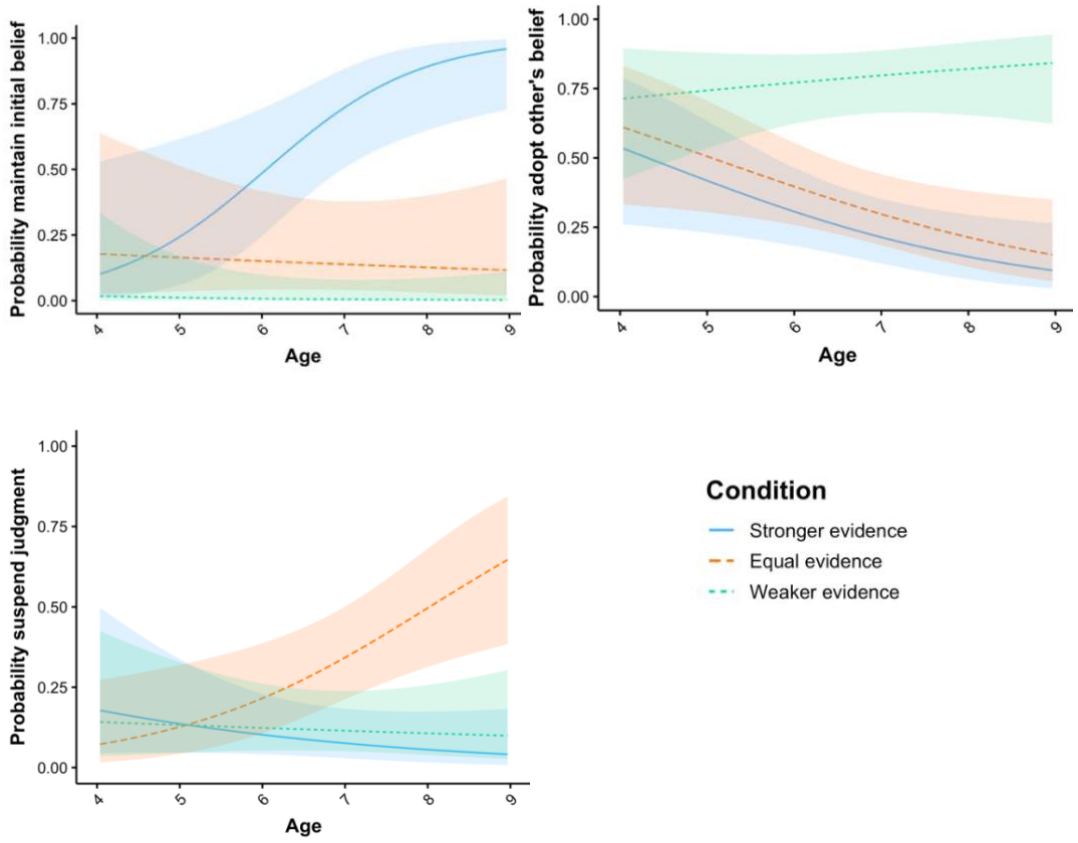
Did participants suspend judgment when their evidence was as good as the evidence of the disagreeing other?

When analyzing participants' responses including age group as a categorical predictor, the full model was highly significant when compared with the null model ($\chi^2(8) = 75.05$, $p <.001$, note that we had to take out the control predictor gender for the model to converge). Specifically, we found a significant interaction be- tween condition and age group ($\chi^2(4) = 15.96$, $p = .003$), while trial number did not have a significant effect ($\chi^2(1) = 3.53$, $p = .06$). Following up on this interaction, we found that in the models for adults and older children, condition was a significant predictor (adults $\chi^2(2) = 81.14$, $p < .001$; older children $\chi^2(2) = 76.81$, $p < .001$). Adults and older children were significantly more likely to suspend judgment in the equal evidence condition compared to the stronger evidence condition (adults $p = .001$, older children $p <.001$), and the weaker evidence condition (adults $p < .001$, older children $p = .033$). In contrast, condition was not a significant predictor in the model for younger children ($\chi^2(2) = 1.17$, $p = .557$). Thus, unlike adults and older children, 4- to 6-year-olds did not show the tendency to suspend judgment in the equal evidence condition; instead, their propensity to suspend judgment was low overall. Consistent with this, within the equal evidence condition, the predominant response of adults and older children was to suspend judgment, while 4- to 6-year-olds were most likely to adopt the other character's belief (see Table 1).

When analyzing children's responses including age as a continuous predictor, we had to take out the con- trol predictor gender and the random slopes for condition. The full-null model comparison was significant ($\chi^2(5) = 27.65$, $p <.001$), revealing a significant interaction between condition and age ($\chi^2(2) = 11.34$, $p = .003$). Specifically, Figure 9 shows that 4- and 5-year-olds' propensity to suspend judgment was similar across conditions. Around age 6, children began to suspend judgment more often in the equal evidence condition than in the other two conditions, and this tendency was more pronounced with increasing age.

Figure 9

Predicted probabilities with which children in Experiment 1 maintained their initial belief, adopted the other's belief, and suspended judgment in the different conditions. Ribbons represent 95% Confidence Intervals.



Discussion

In Experiment 1, we investigated whether children and adults responded reasonably to disagreement. In line with our predictions, we found that 4- to 6-year-olds, 7- to 9-year-olds, and adults maintained their initial belief most often when their own evidence was stronger, and adopted the belief of the other agent most often when their own evidence was weaker. However, we also found interesting developmental differences.

A first central difference was that younger children showed an overall tendency to adopt the belief of the dis- agreeing other. As a group, 4- to 6-year-olds selectively adjusted or maintained their initial belief based on the relative strength of the evidence supporting their own belief versus the belief of the disagreeing other. However, this was mostly driven by the 6-year-olds, as some 5-, and many 4-year-olds adopted the other's belief in both conditions. Thus, a number of our youngest participants may have followed the strategy: "when somebody dis-agrees with you, adopt their belief." This would be in line with previous research showing that although 4- and 5-year-old children maintain their initial belief in the face of conflicting testimonial evidence when their initial belief is clearly based on stronger evidence (Miosga et al.,

2020; Schleihauf et al., 2022; Young et al., 2012), they often adopt the belief of a disagreeing other when it is not perfectly obvious which of the two opposing beliefs is correct (Bridgers et al., 2016; Hagá & Olson, 2017b). The scenarios children encountered in our experiment arguably presented such ambiguity, since participants formed their initial belief based on a limited amount of testimonial evidence. Because of this relative ambiguity, some of our youngest participants might have taken disagreement as a general indicator that they were mistaken, leading them to adopt the belief of the disagreeing other without weighing the evidence supporting each of the opposing beliefs, as older children and adults did.

The second central developmental difference we found in Experiment 1 was that, unlike older children and adults, the group of 4- to 6-year-olds did not engage in metacognitive suspension of judgment more often in the equal evidence condition than in the other two conditions. Across conditions, younger children's likelihood to suspend judgment was low, and it did not differ between conditions. Instead, within the equal evidence condition, 4- to 6-year-olds showed a tendency to adopt the other person's belief rather than to suspend judgment. But can we conclude from Experiment 1 that the ability to suspend judgment in situations of ambiguous evidence does not emerge until around age 7? No. There are at least three different reasons for why 4- to 6-year-olds may have refrained from suspending judgment in the equal evidence condi- tion of Experiment 1.

A first possibility is that 4- to 6-year-olds may be al- together unable to suspend judgment. For example, the ability to suspend judgment might require inhibitory control, which develops only gradually over the preschool- and early school-years (Best & Miller, 2010). Children of this age might also struggle with identifying ambiguous evidence. They may not have realized that in the equal evidence condition of our experiment, the strength of their own evidence and that of the dis- agreeing other were identical. Four- to 6-year-olds might also be unable to suspend judgment because they fail to identify the uncertainty that situations of ambiguous evidence give rise to, due to limited metacognitive abilities (see Kuhn, 2000).

A second possibility is that while 4- to 6-year-olds are generally able to suspend judgment, they may be unwilling to do so, because children of this age are generally reluctant to embrace ambiguity and prefer holding a belief that is based on ambiguous evidence to not holding a belief at all (see Hagá & Olson, 2017b).

Finally, a third possibility is that 4- to 6-year-olds are generally willing and able to suspend judgment when this is operationalized in a different way. In Experiment 1, suspending judgment required participants to make an explicit verbal judgment of their own uncertainty (“I am not sure, I need more information”). Prior work has shown that children of this age often struggle with explicit uncertainty judgments (Butterfield et al., 1988; Coughlin et al., 2015; Hagá & Olson, 2017b; Lipko et al., 2009; Lipowski et al., 2013). Thus, if suspension of judgment were assessed more implicitly, even young children might view this as the most reasonable response in situations of ambiguous evidence. Instead of jumping to the conclusion that the ability to suspend judgment in situations of ambiguous evidence does not emerge until around age 7, we suspended judgment on the issue, and explored it further in Experiment 2.

4.3 Experiment 2

The goal of Experiment 2 was to test one explanation for why 4- to 6-year-olds did not suspend their judgment in a situation of disagreement where the evidence supporting two conflicting beliefs was equally strong. As discussed in the previous section, one possibility is that

in general, 4-to 6-year-olds are able and willing to suspend judgment if it is operationalized in a different way. In order to suspend judgment in Experiment 1, participants had to indicate that they were "not sure" and needed "more information," but they never experienced the actual consequences of suspending judgment; the trial instead ended, and they did not receive any additional information. Previous work from the metacognition literature suggests that when asked to verbally report their own uncertainty, preschoolers struggle with assessing their own knowledge and often overestimate their own abilities (Butterfield et al., 1988; Coughlin et al., 2015; Hagá & Olson, 2017b; Lipko et al., 2009; Lipowski et al., 2013). For example, even when their objective chances of correctly answering a question are low, young children often claim that they are "very sure" that they know the answer – an interesting contrast to their general tendency to adopt others' beliefs (see Hagá & Olson, 2017b). In contrast, studies employing implicit measures suggest that even preschoolers can monitor their uncertainty and draw appropriate consequences from it. For example, preschoolers are more likely to opt out of answering a question (Lyons & Ghetti, 2012) or search for additional information in situations of greater uncertainty (Hembacher et al., 2020).

Thus, even 4- to 6-year-olds might be able to suspend judgment in situations of ambiguous evidence when this is assessed in a more implicit manner. To investigate this possibility, we operationalized suspension of judgment in Experiment 2 as the search for additional information, thus characterizing it as "something one does in order to genuinely inquire" (Friedman, 2017). Specifically, participants were told that if they were not sure, they could ask more people where the bunny went. If participants chose this option, they were exposed to two additional informants, whose statements confirmed either the participant's initial belief or the belief of the disagreeing other. We were interested in whether, under these circumstances, 4- to 6-year-olds would be more likely to suspend judgment in the equal evidence condition than in the other two conditions.

Methods

Participants

Experiment 2 had 104 participants: 54 children aged 4.0–6.11 (27 4-year-olds, 13 5-year-olds, 14 6-year-olds, M_{age} = 5.2years, SD_{age} = 1.02 years, 50% female), and 50 adults as a comparison sample (M_{age} = 30.54, SD_{age} = 12.10, 42% male, 52% female, 4% of participants were non-binary and 2% gender non-conforming). Participants were recruited and compensated as in Experiment 1. Children were predominantly white (37%). 19% of parents chose not to indicate their child's race or ethnicity. Fifteen percent of children were Asian and white, followed by 7% Asian, 6% Latinx, 4% Latinx and white, 4% African or African American, 4% Asian and Pacific Islander. The remaining 4% were Latinx and African American, Latinx and Asian, or Latinx, Asian, and white. In line with our pre-registered exclusion criteria, six additional children were excluded because they responded to the initial belief assessment incorrectly in more than one test trial even after feedback (N = 3), did not complete the study (N = 2), or were older than 6.11 years (N = 1). One additional child was excluded due to parental interference. Data for children were collected between November 2020 and March 2021. Adults were mostly white (60%), followed by Asian (16%), African or African American (10%), Latinx (8%), Latinx and white (4%), and African or African American and American Indian or Alaskan Native (2%). Adult data were collected in December 2020.

A simulation-based power analysis (using the R pack- age simr) – based on the effect of our main predictor, condition, on older children's probability to suspend judgment in Experiment 1 – estimated that this effect had 61.10% power (1000 simulations). Based on this, we calculated that a sample size of 50 participants per age group would give us 91.5% power for the effect of condition on our response variable. We tested four additional children because families were recruited simultaneously and testing appointments had already been scheduled when it became clear that we had reached our target sample size.

Design

The design of Experiment 2 was identical to that of Experiment 1, except that participants were able to ac- quire additional evidence after they decided to suspend judgment. This additional evidence confirmed the participant's initial belief in the stronger evidence condition and the belief of the disagreeing other in the weaker evidence condition. In the equal evidence condition, we counterbalanced between participants whether the additional information confirmed the participant's initial belief or the belief of the disagreeing other.

Procedure

The general procedure for Experiment 2 was similar to that of Experiment 1 but involved some specific changes. First, Experiment 2 contained an additional practice trial. In this practice trial, participants initially asked one informant where the bunny went, but the informant claimed that they had not seen the bunny and thus did not know where the bunny went. The goal of this practice trial was to expose participants to a situation in which suspending judgment was clearly the right thing to do, so that participants would experience what it was like to suspend judgment and ask for additional information before moving onto the test trials. When a participant chose the suspension of judgment response option in this or any of the subsequent test trials, they were exposed to two additional informants who told the participant where the bunny had gone. As in Experiment 1, the informants always provided a consistent response. In the equal evidence condition, we counterbalanced between participants whether the in- formation provided by the two additional informants supported the participant's or the disagreeing other's initial belief.

The main change introduced in Experiment 2 was that the response option representing suspension of judgment was operationalized differently. Instead of being operationalized as “I am not sure. I need more information.”, as in Experiment 1, the suspension of judgment response option was now operationalized as “I am not sure. I would like to ask more people where the bunny went.” If participants had suspended judgment and acquired the additional information, their belief about where the bunny went was assessed for a third time. For example, when a participant had decided to suspend judgment following the disagreement in the stronger evidence condition, the experimenter said (children) /the participant read (adults) “five people told you the bunny went toward the house. And one person told [friend's name] the bunny went toward the bridge. What do you think now? where do you think the bunny went?”. In this example, the response options at were “toward the house”, which rep- resented sticking with one's initial belief, and “toward the bridge”, which represented adopting the friend's belief.

An additional minor change was that in Experiment 2, both children and adults received feedback after the initial belief assessment. This was to make the procedure more similar for both age groups (as adults in Experiment 1 did not receive such feedback). If a participant responded to the initial belief assessment incorrectly for the first time, the test question (“Where do you

think the bunny went?") was repeated. If a participant responded to the initial belief assessment incorrectly for a second time, the participant was prompted with, for example: "One person said the bunny went to the house. So, I think the bunny went to the house. What do you think?". If a participant responded to the initial belief assessment incorrectly for a third time in one of the practice trials, the experimenter said (for children) /the participant read (for adults) "Remember, one person said the bunny went to the house. So, I am going to choose the house.", and the correct response was selected for the participant. In test trials, participants did not receive this third prompt. Instead, they were automatically directed to the subsequent test trial if they responded to the initial belief assessment incorrectly for a third time.

Data analysis

Models were linear mixed regression models (lme4 package in R, Bates et al., 2012) with binomial distributions (logit link). As in Experiment 1, our primary preregistered analyses included the interaction between condition (stronger vs. weaker vs. equal evidence condition) and age group (adults vs. younger children) as the main predictor. However, we could not fit any of the models to our full dataset (i.e., including data from both adults and children) because there was not enough variance in the adult data. We thus report adult data descriptively, and regression results for child data only. The main predictor in these models was condition; when possible, the mod- els included gender (female vs. male) and trial number (1–3, z-transformed) as control predictors, and a ran- dom intercept for subject. As secondary preregistered analyses, we investigated the effect of the interaction between condition and age as a continuous predictor (in years and months) on children's responses, as in Experiment 1.

Results

As for Experiment 1, we organize this section according to our predictions about how a reasonable person should respond to disagreement. We analyze whether (1) participants maintained their initial belief more often in the stronger evidence condition, (2) adopted the belief of the disagreeing other more often in the weaker evidence condition, and (3) suspended judgment more often in the equal evidence condition than in the respective other two conditions. Again, our analyses are based on trials in which participants responded correctly to the initial belief assessment (after feedback). In total, participants failed to respond correctly to the initial belief assessment on five trials (adults: 1 trial in the weaker evidence condition; children: 1 trial in the stronger evidence condition, 3 trials in the weaker evidence condition).

Did participants maintain their initial belief when their own evidence was stronger?

In the stronger evidence condition, adults maintained their initial belief 82% of the time, while they did so only 10% of the time in the equal evidence condition and never in the weaker evidence condition. In the model including the effect of condition on children's responses, condition was a significant predictor ($\chi^2(2) = 8.65$, $p = .013$). There was also a significant effect of trial number ($\chi^2(1) = 7.46$, $p = .006$), such that 4- to 6-year-olds were more likely to maintain their initial belief in the third test trial compared to the first or second test trial. There was no significant effect of gender ($\chi^2(1) = 2.39$, $p = .122$). Thus, as a group, 4- to 6-year-olds were more likely to maintain their initial belief in the stronger evidence condition than in the other two conditions, as predicted. However, post hoc tests were not significant ($p = .094$ for the stronger vs. equal comparison; $p = .081$ for the stronger vs. weaker comparison). When comparing

participants' responses within the stronger evidence condition, both adults and children were most likely to maintain their initial belief. However, 4- to 6-year-olds were almost as likely to adopt the other's belief (see Table 1). For the model including the interaction between condition and age as a continuous predictor, we had to take out the control predictor gender and the random slopes. Compared to the null model, the full model was significant ($\chi^2(5) = 20.68$, $p = .001$), revealing a reliable effect of the interaction between condition and age ($\chi^2(2) = 8.71$, $p = .013$). The tendency to maintain one's initial belief more often in the stronger evidence condition than in the other two conditions first emerged around age 5, and became increasingly pronounced with age (Figure 10).

Did participants adopt the other's belief when their own evidence was weaker?

Eighty-two percent of adults adopted the other's belief in the weaker evidence condition. In the equal evidence condition and the stronger evidence condition, only 4% and 1% of adults adopted the friend's belief, respectively. In the model predicting children's propensity to adopt the friend's belief, condition was a significant predictor ($\chi^2(2) = 11.15$, $p = .004$). There was no significant effect of the control predictor trial number ($\chi^2(1) = 0.37$, $p = .55$) or gender ($\chi^2(1) = 2.16$, $p = .14$). Specifically, as predicted, children adopted the friend's belief more often in the weaker evidence condition than in the stronger evidence condition, and in the weaker evidence condition than in the equal evidence condition. Again, post hoc tests did not reach significance ($p = .075$ for the weaker vs. stronger comparison; $p = .063$ for the stronger vs. equal comparison). When comparing data within the stronger evidence condition, adopting the other's belief was the predominant response in both age groups, although the difference between adopting the other's belief and the other two response options was more pronounced in adults (Table 1).

For the model including age as a continuous predictor, we had to take out gender and the random slopes. The full model was significant when compared to the null model ($\chi^2(5) = 35.08$, $p<.001$), revealing a significant interaction of condition and age ($\chi^2(2) = 21.60$, $p<.001$). As Figure 10 shows, children were more likely to adopt the other's belief in the weaker evidence condition and less likely to do so in the other two conditions with increasing age.

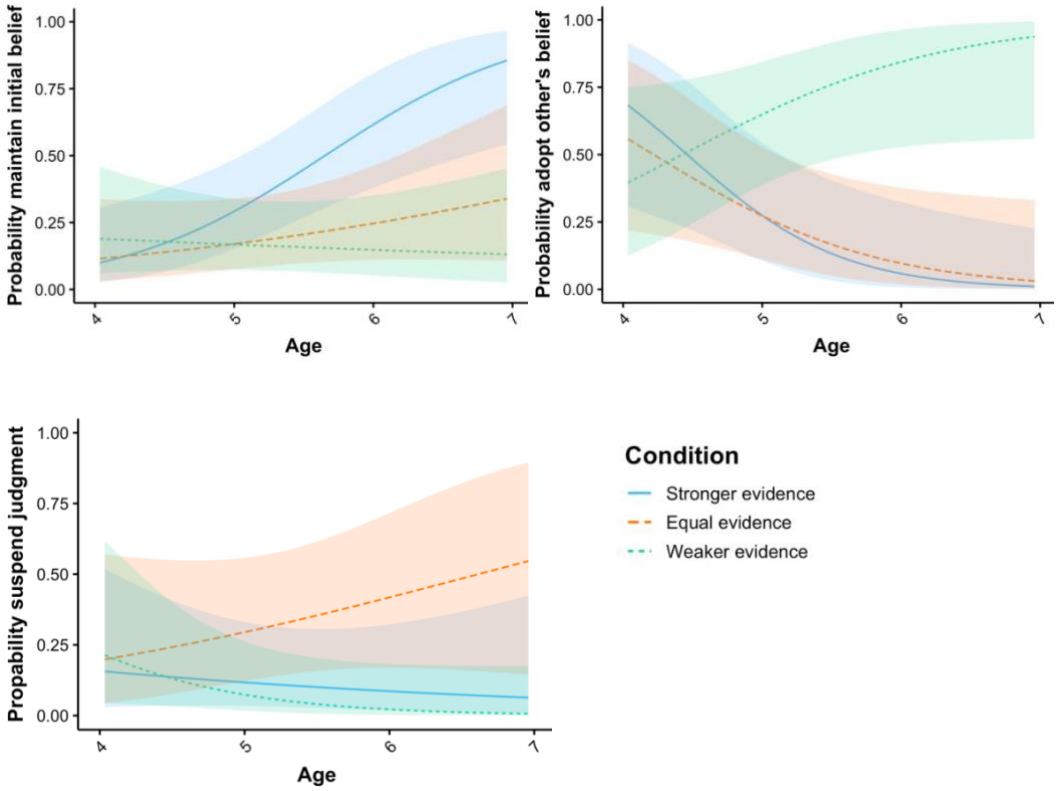
Did participants suspend judgment when their evidence was as good as the evidence of the disagreeing other?

Seventy-six percent of adults suspended judgment in the equal evidence condition. In contrast, only 14% did so in the stronger evidence condition, and 16% in the weaker evidence condition. Unlike in Experiment 1, condition was a significant predictor of 4- to 6-year-olds' responses ($\chi^2(2) = 13.94$, $p<.001$). There was no significant effect of trial number ($\chi^2(1) = 0.75$, $p = .387$) or gender ($\chi^2(1) = 0.77$, $p = .380$). Specifically, 4- to 6-year-olds were significantly more likely to suspend judgment in the equal evidence condition than they were in the stronger evidence condition ($p <.001$), and in the weaker evidence condition ($p = .003$). After suspending judgment and acquiring the additional information, the majority of participants from both age groups converged onto the belief that was supported by more evidence overall (see Supplementary Material). Within the equal evidence condition, the suspension of judgment option was the modal response in both age groups (see Table 1). However, while suspension of judgment was the clear, majority response among adults (85%), it remained a minority response among 4- to 6-year-olds (41%).

For the model including age as a continuous predictor, we had to take out the control predictor gender and the random slopes. The model was significant when compared with the null model ($\chi^2(5) = 17.64$, $p = .003$). Specifically, there was a significant effect of the condition * age interaction ($\chi^2(2) = 7.36$, $p = .025$). Children were increasingly likely to suspend judgment in the equal evidence condition with age, and less likely to do so in the other two conditions (Figure 10).

Figure 10

Predicted probabilities with which children in Experiment 2 maintained their initial belief, adopted the other's belief, and suspended judgment in the different conditions. Ribbons represent 95% Confidence Intervals.



Discussion

In Experiment 2, we focused specifically on 4- to 6-year-olds' developing ability to suspend judgment in situations of ambiguous evidence – a metacognitive ability which has been discussed as a key competence in Philosophy (Frances, 2014; Friedman, 2017) and Education (Haney, 1964). Our results show that with increasing age, 4- to 6-year-olds are more willing and able to suspend judgment in situations of disagreement if it is assessed in an implicit manner. Specifically, rather than assessing suspension of judgment as an explicit, metacognitive judgment about participants' uncertainty (as in Experiment 1), suspension of judgment in Experiment 2 was operationalized via the search for additional information.

When comparing 4- to 6-year-olds' propensity to suspend judgment in Experiment 1 and Experiment 2 (see supplementary materials), we found that this procedural change led to a slight overall increase in 4- to 6-year-olds' propensity to suspend judgment (i.e., across conditions). That is, it seems that searching for information was overall more attractive for children compared to making explicit judgments. Although older children were not included in Experiment 2, we speculate that the different operationalization would have had a similar impact on older children, as well. Importantly, however, the increase in suspension of judgment responses among 4- to 6-year-olds was most dramatic in the equal evidence condition: as a group, 4- to 6-year-olds in Experiment 2 suspended judgment significantly more often in the equal evidence condition than in the other two conditions – this had not been the case in Experiment 1. Thus, our results suggest that assessing suspension of judgment more implicitly made it easier for 4- to 6-year-olds to respond rationally to the ambiguous evidence in the equal evidence condition. These findings are in line with previous work showing that young children show evidence of implicitly monitoring their uncertainty in their information seeking behavior (Coughlin et al., 2015; Hembacher et al., 2020; Lapidow et al., 2022; Perner, 2012).

While 4- to 6-year-olds in Experiment 2 were more likely to suspend judgment in the equal evidence condition than in the other two conditions, and more likely to do so than in Experiment 1, we note two caveats. First, the differences between 4- and 6-year-olds' propensity to suspend judgment in the equal evidence condition compared to the other two conditions were far less pronounced than in adults. And second, within the equal evidence condition, only 41% of all 4- to 6-year-olds suspended judgment. Thus, the modal response for 4- to 6-year-olds within the equal evidence condition was suspension of judgment (41%). Still, the majority of 4- to 6-year-olds did not suspend judgment, and instead either maintained their initial judgment (24%) or adopted the friend's belief (35%). What might explain this response pattern? One possibility is that even with the more implicit operationalization, many 4- to 6-year-olds may have failed to realize that there was equal evidence for the two competing claims – and/or that this meant that there was uncertainty about the bunny's whereabouts – and thus that additional evidence was needed to resolve the disagreement. Alternatively, these children may have understood that there was uncertainty about the bunny's location but did not opt to suspend judgment because they preferred holding a belief based on ambiguous evidence to not holding a belief at all. Future research could investigate more systematically which of these (or other) factors best explains 4- to 6-year-olds' limited ability to suspend judgment.

In addition to showing that 4- to 6-year-olds were more likely to suspend judgment in the equal evidence condition than in the other two conditions, Experiment 2 also replicated some of the key findings of Experiment 1. Four- to 6-year-olds, as well as adults, responded reasonably to disagreement in that they were more likely to maintain their initial belief when their own evidence was stronger (compared to weaker or equal), and adopt the belief of the disagreeing other when their own evidence was weaker (compared to stronger or equal). However, unlike in Experiment 1, the post hoc tests comparing the frequency of maintaining one's initial belief and adopting the friend's belief across conditions were not significant in children. A look at Table 1 suggests that children in Experiment 2 were more drawn to the sus- pending judgment option across conditions (perhaps due to how we presented this option), which may have caused the contrasts between the response options to be less pronounced overall compared to in Experiment 1. Yet, taken together, the results of Experiment 2 suggest that as a group, even 4- to 6-year-olds demonstrate intellectual humility in responding reasonably to disagreement.

4.4 General Discussion

When disagreeing with others on issues that have a single, objectively correct answer, a reasonable person should adjust their beliefs based on which of the two opposing beliefs is supported by stronger evidence, and sus- pend judgment if both beliefs are supported by equally strong evidence (Frances, 2014). We investigated the development of this ability across two experiments. In Experiment 1, participants of all age groups responded reasonably in maintaining their initial belief when their own evidence was stronger (stronger evidence condition), and adopting the other's belief when their own evidence was weaker (weaker evidence condition). However, one central developmental difference was that 4- to 6-year-olds did not reliably suspend judgment when their evidence was as good as that of the disagreeing other (equal evidence condition). In contrast, in Experiment 2, where suspension of judgment was assessed more implicitly – via the search for additional information – 4- to 6-year-olds were significantly more likely to sus- pend judgment in the equal evidence condition than in the other two conditions (although 4-year-olds did still not suspend judgment reliably). This finding is in line with other research showing evidence of implicit uncertainty monitoring in children's information seeking behavior (Coughlin et al., 2015; Hembacher et al., 2020; Lapidow et al., 2022; see Perner, 2012, for discussion of whether implicit measures of uncertainty are valid measures of metacognition).

Together, our studies demonstrate the presence of social reasoning skills and intellectual humility from at least age 5 onwards. In showing that children begin to respond reasonably to disagreement from early on, our findings align with recent accounts in cognitive science that have emphasized the social dimensions of reasoning as a whole (Heyes, 2012; Köymen & Tomasello, 2020; O'Madagain & Tomasello, 2021; Tomasello, 2020), and metacognitive reasoning, in particular ((e.g., Dutilh Novaes, 2018; 2020; Mascaro & Sperber, 2009; Heyes, 2018; Tomasello, 2019; Kuhn, 2019; Nagel, 2015; O'Madagain, 2019). For such accounts, the ability to respond to disagreement by revising one's belief in line with evidence – or suspend judgment if additional evidence is required – constitutes a key milestone on the way to become a rational reasoner. Such responses to disagreement require that a reasoner understand that others have perspectives that differ from their own, and integrate these perspectives with the objective state of the world; in our case by comparing the evidence supporting each of the two opposing beliefs (see Tomasello, 2020). In addition to illustrating the development of a crucial metacognitive skill, our research also provides an important practical insight: it suggests that from early in life, children possess the cognitive prerequisites needed to develop into autonomous, open-minded citizens who can contribute to a meaningful public discourse. A meaningful public discourse, in turn, constitutes the backbone of any healthy democracy.

Although as groups, adults, older children, and younger children responded reasonably to disagreement in the current studies, the present findings also demonstrate a significant age-related "increase in reasonableness." First, the analyses including age as a continuous predictor revealed that 4-year-olds did not actually dis- criminate between conditions. They were neither more likely to maintain their initial belief when their own evidence was stronger, nor to suspend judgment when their evidence was as strong as that of the disagreeing other. Instead, 4-year-olds almost always adopted the other's belief. Second, in both experiments, children discriminated more strongly between conditions with increasing age, and adults discriminated most strongly. What might explain these age differences? One option is that with increasing age, children may

have simply found it easier to pay attention to the experiment. Another, more interesting possibility is that the ability to respond reasonably to disagreement undergoes significant development throughout childhood. If the latter is the case, what factors are responsible for this development?

One factor that may have played a role in our studies may be children's developing ability to reliably compare how many informants support a specific claim (see Morgan et al., 2015). Importantly, experimentally manipulating how many informants support a claim is only one of many possible ways of manipulating the strength of evidence for a belief. In addition to manipulating the strength of evidence (i.e., the quantity of information), future research could also investigate how children respond to disagreements when opposing beliefs are supported by evidence of different quality, such as stronger or weaker reasons (see, e.g., Koenig, 2012; Köymen & Tomasello, 2018; Mercier et al., 2018).

A capacity that may specifically affect children's developing ability to suspend judgment – and that has been argued to share crucial features with metacognition (see Roebers, 2017) – is inhibitory control (see e.g., Wiebe et al., 2012). As their inhibitory control increases, children might increasingly avoid "jumping to conclusions," and understand that it is rational to suspend judgment about a given belief until one has acquired sufficient evidence (see Hagá & Olson, 2017a, 2017b; Crawford, 2022). We tested for a potential relation between participants' inhibitory control skills (assessed via a go/no– go task, e.g., Wiebe et al., 2012), and their propensity to suspend judgment in our Experiment 2 (details on this analysis are reported in the Supplementary Materials). Although we did not find a relation between participants' inhibitory control skills here, we see this as an interesting avenue for future investigation.

To conclude our discussion of factors that might lead to age-related increases in reasonableness, we want to highlight that what is considered "reasonable" might itself change across development. What constitutes a reasonable response for adults does not necessarily constitute a reasonable response for children. Specifically, we observed across both experiments that younger children showed a general tendency to adopt a disagreeing other's belief. While this is not in line with the here de- fined criteria for reasonableness, it is important to note that this strategy need not be considered "unreasonable." Young children's knowledge about the world is still limited. They frequently experience that they themselves are wrong about things and that others are right, even in situations where the evidence supporting their belief may have seemed strong. For someone who knows little and is used to interacting with knowledgeable adults, being open toward adopting others' conflicting beliefs arguably constitutes an adaptive learning strategy.

Importantly, possessing the ability to respond reasonably to disagreement does not necessarily mean that people actually make use of that ability (see Kuhn, 2022). Individuals often respond to disagreement in ways that one might perceive as "unreasonable" (at least from an epistemic perspective). People can be hesitant to give up their beliefs, even in light of strong evidence for an alternative view. Prior research with adults shows that this is often the case when an individual's belief carries personal importance, for example, because it is closely connected to their identity (e.g., Kahan, 2012; Lewandowsky & Oberauer, 2016). One explanation for this is that letting go of a belief that is closely connected to one's identity can carry immense social and emotional costs, such as being excluded from one's social group or losing one's sense of belonging. Against this background, it would not be surprising if children – like adults – would forgo pursuing the epistemic goal of "taking on the belief that is best supported by evidence" in

favor of protecting their identity (see Kelly, 2003). It is up to future research to show this empirically.

More generally, future studies should focus on how children balance the relative costs and benefits of different responses to disagreement. With regard to suspension of judgment, one study found that, in contrast to older children and adults, young children who overheard speakers talking about ambiguous objects did not favor "intellectually humble" speakers (who acknowledged that they might be wrong), compared to diffident speakers or intellectually arrogant speakers (Hagá & Olson, 2017a). In other contexts, even older children and adults might negatively evaluate individuals who suspend judgment, because these individuals could be perceived as lacking in self-confidence or as insecure. An interesting avenue for future research would thus be to identify the contextual factors that contribute to a potential shift from perceiving suspension of judgment as "humble" and desirable to perceiving it as hesitant and fickle. With regard to the relative costs and benefits of adopting other people's beliefs, it is well known that both children (Haun & Tomasello, 2011) and adults (e.g., Asch, 1956) are more likely to respond incorrectly to a simple perceptual question after a group of confederates has responded incorrectly. In these situations, participants' responses do not necessarily represent their true beliefs – when asked individually, participants almost always make the correct perceptual judgment. Nevertheless, this research indicates that conformity- or authority-related concerns might lead people to adopt others' beliefs even if these beliefs are not supported by stronger evidence. In some communities, respect for authorities plays a more important role than in others, perhaps because the community values integration and social responsibility to the group (Greenfield et al., 2000). This raises the possibility that, in some cultural contexts, maintaining one's initial belief when disagreeing with older or other higher-status members of the community might simply not be perceived as appropriate, even if the evidence supporting one's initial belief is stronger. Moreover, parenting styles might affect the extent to which children feel encouraged to speak up to defend their own beliefs as opposed to adopting their parents' beliefs without questioning (e.g., Baumrind, 1971).

In conclusion, the current experiments present evidence that children from at least age 5 onwards respond reasonably when being confronted with the opposing belief of a disagreeing other. Thus, the metacognitive ability to reflect on one's own knowledge and adjust one's beliefs according to evidence – or suspend judgment if additional evidence is required – is present from at least the preschool years. However, our results also show that individuals' metacognitive responses to disagreement improve with age. Going forward, research should study how children respond to disagreement in a variety of social environments, since we know from research with adults that their responses to disagreement can be heavily influenced by a variety of social and contextual factors.

Chapter 5: Discussion

In 1922, the philosopher, psychologist, and educational reformer John Dewey lauded disagreement as "a sine qua non of reflection and ingenuity". Like him, many other influential thinkers have pointed to the benefits of disagreement for science, society, and the individual (e.g., Arendt, 1958; Habermas, 1962, 1997; Piaget, 1952). Despite these theoretical contributions, the specific individual psychological benefits of disagreement have long remained unclear. Addressing this gap, the current dissertation systematically investigated a specific positive consequence of disagreement for individual cognition: its potential to enhance young children's metacognitive reflection. The effects of disagreement on young children's metacognition were explored across three chapters, each focusing on a distinct, frequently studied dimension of metacognition: reason-giving (Chapter 2), confidence ratings (Chapter 3), and rational belief revision (Chapter 4).

Chapter 2 demonstrated that children from three diverse cultural backgrounds were more likely to provide reasons for their beliefs when disagreeing, as compared to agreeing with their partner about where to find a reward – thus using their reasons selectively to make effective joint decisions. Chapter 3 revealed that experiencing disagreement (versus agreement) reduced young children's overconfidence, and also increased their motivation to search for the correct answer. Finally, Chapter 4 showed that disagreement prompted children to consider the reliability of the evidence supporting their beliefs, and flexibly revise these beliefs or suspend judgment until they had acquired additional evidence. Together, these findings-confirm the notion that experiencing disagreement prompts young children's metacognitive awareness – an insight that holds significant theoretical and practical implications.

Theoretically, the current findings are significant because they bridge the gap between two influential perspectives on the role of social interaction in cognitive development. On one side are constructivist learning theories, which have long suggested that experiencing disagreement can generate learning and intellectual growth, but have not specified the cognitive processes through which disagreement might lead to enhanced learning outcomes (Dewey, 1929; Moshman, 1995; Piaget, 1952; Vygotsky, 1987). On the other side are cultural evolutionary accounts of metacognition, which have recently argued that human metacognition is culturally acquired and originates in conversation-based social interactions (Dutilh Novaes, 2018; 2020; Heyes, 2018; Kuhn, 2019). However, these latter accounts have not pinpointed the specific types of social interaction most conducive to driving metacognitive development. The current work unites these two influential theoretical accounts by showing that a specific form of social interaction – disagreement – facilitates a particular cognitive process that has been shown to enhance learning: metacognition. In doing so, this study takes an important first step toward testing the broader hypothesis that disagreement indeed plays a key causal role in the development of children's metacognitive abilities.

In addition to beginning to illuminate the social origins of metacognition, the current work also sheds light on its social functions (see Heyes et al., 2021; Nagel, 2015). Across five studies, it shows that within the realm of social disagreement, even young children demonstrate remarkable metacognitive capacities (see also Baer et al., submitted; Mahr et al., 2021, for similar findings). In doing so, they constitute a significant departure from much of the earlier work on metacognitive development, which has typically assumed that the primary goal of metacognition is to serve individual learning and decision-making, and thus evaluated children's metacognitive abilities in solitary settings (e.g., Flavell, 2000). In such settings, children often

exhibit significant limitations in their metacognitive abilities until at least late childhood (e.g., Gopnik & Graf, 1988; O'Neill & Gopnik, 1991; van Loon et al., 2017). This suggests that past developmental work may have underestimated the metacognitive abilities of young children by assessing them in isolated contexts.

Besides their theoretical contributions, the current findings also have significant practical implications. First, they suggest that disagreement could serve as a valuable tool in early education, by helping young children improve their metacognitive awareness and therefore their learning. Although disagreement is already incorporated into various educational approaches, including problem-based learning (Wood, 2003), project-based learning (Kokotsaki et al., 2016), and cooperative learning (Johnson & Johnson, 2009), these approaches typically target older learners – perhaps because educationalists typically assume that young children may not yet benefit from working through disagreements.

The present findings challenge this assumption, suggesting instead that disagreement could be particularly advantageous during the early years. Thus, systematically exposing younger children to disagreement within their learning environments – such as by pairing children with differing intuitive theories or prior assumptions to collaborate on a science problem – could be a promising approach to fostering metacognitive reflection and enhancing learning. Interestingly – although there is of course significant variation between schools – this approach seems to contrast with the emphasis commonly found in many educational institutions. Particularly in pre-K and Kindergarten, there often seems to be a stronger focus on harmony (see, e.g., Hanish et al., 2016), and children might even be taught to avoid disagreements and conflicts with their peers. In contrast, the current findings highlight the importance of encouraging respectful disagreement even among the very youngest students, as this could be crucial for promoting their critical thinking, knowledge generation, and overall learning.

Beyond their potential educational advantages, the current findings could also offer important insights for interventions aimed at fostering essential citizenship skills that could contribute to a healthy democracy. As mentioned in Chapter 1, numerous scholars have underscored the significance of disagreement for scientific and societal progress (Arendt, 1958; Dewey, 1922; Habermas, 1997; Sen, 2003). However, disagreement can only fulfill its role in advancing science and societies if individuals respond to it in rational and reflective ways. The research presented in this dissertation shows that the cognitive prerequisites for such responses are present from early in life. This insight could be an important starting point for designing interventions aimed at educating individuals of all ages to become autonomous, open-minded citizens.

Although the current findings hold substantial theoretical and practical significance, there remains ample room for additional exploration regarding the relation between disagreement and metacognition. In the remainder of this Chapter, I will explore three key directions for further inquiry. First, what is the underlying psychological mechanisms through which disagreement might be generating its beneficial effects on metacognition, and how could it be tested (Section 5.1)? Second, what are the next steps in investigating the broader idea that disagreement constitutes the key causal driver of metacognitive development (Section 5.2)? And third, what are the boundary conditions of the benefits of disagreement? That is, when does disagreement foster metacognitive reflection and learning, and when might it be less effective – or even lead individuals to become less receptive to alternative perspectives than they would otherwise be (Section 5.3)?

5.1 What is the underlying psychological mechanism?

The work presented in this dissertation has successfully established that the experience of disagreement positively affects young children's metacognition. However, it does not directly speak to which underlying psychological mechanism might be generating these positive effects. In the current section, I discuss what might be going on in children's heads when they encounter disagreement, and how this might generate metacognitive change.

Recognizing the presence of alternative perspectives

The first step toward benefiting metacognitively from disagreement is arguably to recognize and represent the presence of a hypothesis or perspective that differs from one's own. This was already acknowledged by Piaget (1952), who suggested that encountering cognitive conflict fosters children's learning and intellectual growth. However, to Piaget (1952), it was not particularly important how this conflict came about – i.e, whether it was physical (e.g., encountering new evidence contradicting one's beliefs) or social (i.e., disagreement) in nature. In contrast, a key idea underlying the current work is that young children may uniquely benefit from experiencing conflict in the form of social disagreement. As explained in Chapter 1, the idea is that disagreements might make alternative perspectives particularly salient by embodying them within another individual (see Kuhn, 2019). This, in turn, might make it easier for young children to represent the alternative perspective *as* an alternative perspective – a task that young children generally often struggle with (see Leahy et al., 2020).

As mentioned in Chapter 1, there is already some preliminary evidence that supports this idea (Doise et al., 1975; O'Madagain et al., 2022). Specifically, children have demonstrated more advanced metacognitive abilities (O'Madagain et al., 2022) and greater learning (Doise et al., 1975) after experiencing disagreement than after experiencing a physical, intrapersonal conflict (e.g., viewing the same object from different visual angles). However, additional research is needed to test this component of the proposed underlying psychological mechanism in more detail. One aspect that future research should pay more attention to are possible developmental changes in children's ability to recognize alternative hypotheses in social versus non-social contexts. One interesting possibility is that younger children might be more likely to recognize alternative perspectives – and thus benefit more metacognitively – when these are presented to them in the form of a disagreement, while older children might benefit from social and non-social conflicts equally. In addition to further behavioral work, a promising avenue for future research lies in computational work. To this end, a formal framework that illustrates how the experience of disagreement can increase the likelihood of representing alternative hypotheses during learning and reasoning is currently being developed by Langenhoff, Thompson and colleagues (in prep).

Resolving the conflict between the two perspectives

However, recognizing the presence of an alternative perspective is only the first step. For disagreement to lead to positive metacognitive change, the individual experiencing the disagreement must additionally recognize and attempt to reconcile the *conflict* between these two perspectives, by seeking additional clarifying information and / or adjusting their beliefs in

reasonable ways (Kuhn, 2022; Tomasello, 2019). Thus, the second component of the psychological mechanism underlying the effects of disagreement on metacognition might be something like the realization that the two perspectives conflict with one another, and that additional evidence is required to determine which of them aligns more closely with the truth (Christensen, 2009; Frances, 2014; Friedman, 2013; Lackey, 2010; O'Madagain & Tomasello, 2021).

In the studies presented across this dissertation, children clearly seem to have made this inference. For instance, in the studies in Chapter 3, children who experienced disagreement notably reduced their confidence regarding which types of toys activated the machine. Moreover, they also attempted to obtain the relevant empirical evidence that would allow them to distinguish between the competing hypotheses and reduce the ambiguity of the situation. Interestingly, some recent research shows that from around age 10, children also make the reverse inference, and reason that an ambiguous stimulus is more likely to have provoked a disagreement than an unambiguous stimulus (Amemiya, Walker, & Heyman, 2021; Amemiya et al., 2023; Yang et al., 2023).

Importantly, however, whether there is a conflict between alternative perspectives depends on the type of disagreement being experienced. The current dissertation focuses specifically on factual disagreements, where a clear ground truth usually exists. In such disagreements, the alternative perspectives are mutually exclusive (e.g., Zeman, 2019). However, as highlighted in Chapter 4, people also frequently disagree on subjective issues, such as preferences, values, opinions, moral beliefs, or aesthetics. These subjective disagreements lack a clear ground truth, allowing both perspectives to be equally valid and coexist without needing integration (see, e.g., Foushee & Srinivasan, 2017; Heiphetz et al., 2013; Kuhn et al., 2002).

An interesting avenue for future research would be to investigate children's understanding of these differences between factual and subjective disagreements. Some prior studies have looked into this by telling children about both factual and subjective disagreements and asking them whether only one or both participants of the disagreement could be right (e.g., Banerjee et al., 2007; Danniels & Perlmann, 2021; Heiphetz et al., 2013; Yang et al., in press; Wainryb et al., 2004). These studies show that starting around age 5 – and increasingly so with increasing age – children tend to judge that only one participant of a *factual* disagreement can be right, while they are more likely to acknowledge that both participants of a *subjective* disagreement can be right. However, the paradigms used in these studies are usually relatively complex and heavily reliant on language. It seems plausible that children might demonstrate an earlier understanding of the differences between factual and subjective disagreements if this was assessed in more intuitive, behavioral ways.

5.2 Disagreement as a key driver of metacognitive development

The findings in this dissertation show that exposure to disagreement prompts young children to engage in metacognitive reflection in the moment of the disagreement itself. However, the broader idea underlying the current work is that disagreement has a more significant and lasting impact than merely inducing a momentary "reflective mindset" (see Kuhn, 2022). Specifically, the claim is that experiences of disagreement might constitute the *key driver* of children's metacognitive development. This claim entails two key points. First, that children's general metacognitive development is causally driven by their (everyday life) experiences of disagreement. And second, that disagreement has a greater impact on metacognitive

development than other factors, such as direct metacognitive instruction from adults. In what follows, I will outline how each of these hypotheses could be tested in future research.

Long-term causal influences

How could one investigate whether disagreement exerts a general causal influence on metacognitive development? One initial step in testing this broader claim would be to establish whether the benefits of disagreement found in the current studies would extend to different situations and contexts. For example, would young children who encounter disagreement after forming a belief based on ambiguous evidence display reduced overconfidence not only regarding the topic of the disagreement itself (as illustrated in Chapter 3), but also in a novel situation involving ambiguous evidence? I am not aware of any studies that have examined such long-term consequences of disagreement on metacognitive reasoning. However, one study has shown that experiencing disagreement regarding their moral beliefs in one situation enhances children's moral reasoning in a different context (Li & Tomasello, 2022). These findings are promising and hint at the potential for similarly generalizing effects within the domain of metacognition.

Additional compelling evidence could be derived from studies examining whether children who, on average, experience more disagreement in their daily lives show accelerated metacognitive development. One factor that might systematically influence how much disagreement young children naturally experience is how much time they spend with same-aged peers, as children might be more comfortable disagreeing with a reasoning partner who is at their eye-level and epistemic level, compared to a seemingly all-knowing adult (see Piaget, 1952). How much time children spend with their peers, in turn, can vary notably across cultural contexts. In particular, in many Western societies, young children typically spend the majority of their time with their immediate core family, which often consists only of their parents and maybe sibling(s); particularly before they begin formal schooling (see Hrdy, 2009). In contrast, in many other cultural contexts, even young children often spend the majority of their time surrounded by other children (Kinoti, 2010; Whiting, 1996). It would be interesting to systematically investigate whether children growing up in the latter cultural environments do indeed encounter disagreements more regularly and thus demonstrate better metacognitive awareness from earlier in life than their Western counterparts.

Another factor that might influence young children's natural exposure to disagreement are parenting styles. For example, authoritarian parents generally expect obedience from their children without providing many explanations. In contrast, authoritative parents set clear rules and expectations, but encourage independent thinking and open communication (see, e.g., Baumrindt, 1971; Bornstein, 2013). As a consequence, disagreement may be discouraged in authoritarian households, while authoritative parents might encourage their children to engage in healthy and constructive disagreements. This, in turn, could lead to differences in children's metacognitive development.

Finally, for a more controlled approach, longitudinal studies could provide additional valuable insights. For example, one could assign children to two groups, for example within the context of a science class. In one group, children are systematically exposed to peer disagreement, while in the other, children encounter more teacher-centered interactions. The metacognitive and learning outcomes of these groups could then be traced and compared over an

extended period of time (see Kuhn & Crowell, 2011, for a similar, though less controlled, approach with adolescents). Taken together, research investigating the generalizability and the long-term effects of disagreement on young children's metacognition will be a crucial next step in determining whether disagreement constitutes a key driver of metacognitive development, but also in establishing the potential for implementing disagreement-based educational interventions.

Unique influences

To further investigate the second key claim – that disagreement uniquely impacts metacognitive development compared to other factors – future research should compare the influence of disagreement with that of children's non-social experiences, as discussed in Section 5.1. Additionally, it is crucial to examine whether disagreement offers unique benefits for metacognitive development compared to other forms of social interactions. As a first step in this direction, the studies reported in Chapters 2 and 3 directly compared disagreement to one other social-communicative context: agreement. In contrast to disagreement, agreement did not yield any discernible metacognitive benefits; if anything, agreement led to children be even *less* aware of their cognitive limitations (e.g., children's overconfidence increased; see Chapter 3).

While this tells us that children's metacognitive awareness is not facilitated by *any* kind of social interaction, the effects of disagreement should also be compared to those of explicit metacognitive instruction from adults. This is particularly relevant because some proponents of cultural evolutionary accounts have argued that it is primarily through such explicit metacognitive guidance that children learn how to interpret metacognitive feelings (e.g., "this is what it feels like when you are not so sure"; Heyes, 2018; see also Efklides, 2006). While I am not aware of any research that has investigated this question directly, there is work showing that labeling can facilitate children's understanding of *other* feelings, such as happiness, anger, or sadness (e.g., Price et al., 2022).

These findings make it plausible that mastering metacognitive language can contribute to fostering metacognitive development (see also Lohmann & Tomasello, 2003). However, it is difficult to see how the flexible, context-specific understanding characteristic of genuine metacognitive competence would result from metacognitive instruction alone. For instance, it seems unlikely that adult-instruction alone could lead children to understand the intricate differences between what it means to feel "really sure" versus "kind of sure" versus "a little sure" (let alone the many intricate "confidence steps" that lie in between these commonly used anchor points). A more plausible scenario is that explicit instruction and peer disagreement work together to shape metacognitive development. Initially, a child might learn to label a metacognitive state through explicit instruction from an adult. However, it seems crucial that the child then further refines this understanding by practicing its use during disagreements with peers. Future research should explore this possibility, paying careful attention to whether any potential differences are qualitative or more quantitative in nature (i.e., disagreements simply happen to be the contexts in which young children most often learn that others can have conflicting beliefs).

5.3 Which factors modulate the effects of disagreement on metacognition?

I began this dissertation by noting that disagreement has been widely recognized for its beneficial consequences, both at the collective, as well as at the individual level. However, disagreements can undeniably also lead to significant negative consequences: they can create

tension, strain relationships, fuel societal division, and even result in violent conflicts. Therefore, from both a theoretical and practical standpoint, it is important to understand exactly which types of disagreements are beneficial and should be fostered, and which could be detrimental and should be avoided. In the current section, I will begin to address this question, focusing specifically on factors that can modulate the relation between disagreement and children's metacognition.

But before diving into this topic, it is important to clarify some key points. Thus far, this dissertation has primarily focused on children's (developing) metacognitive competence. Yet, research (and everyday experiences) with adults show that possessing the ability to respond to disagreement via metacognitive reflection does not necessarily mean that one is also willing to *use* this ability (e.g., Kahan et al., 2017).

Disagreements are fundamentally social contexts. In such contexts, individuals are motivated not only by their epistemic goals of seeking truth or developing accurate worldviews – which foster metacognitive reflection – but also by various social considerations such as the motivation to establish a positive personal reputation, maintain a coherent sense of one's self, conform to the norms of one's ingroup, or signal social status (see Yoon et al., 2018). When an individual experiences a disagreement in which their epistemic and their social goals conflict, and the individual weighs the social goal more heavily than the epistemic goal, the metacognitive benefits of disagreement might be stifled. Importantly, although weighing a social goal more heavily than an epistemic goal could be viewed as "irrational" from an epistemic perspective, such behavior may be *socially* very rational, particularly when the costs of abandoning one's social goals would be high (Kelly, 2003).

Against this background, an important question is what determines whether an individual prioritizes epistemic or social goals in a given disagreement context. On the one hand, this might depend on development. Young children are known to be particularly curious, and more inherently motivated to obtain information, than older children and adults (e.g., Liquin & Gopnik, 2022). Moreover, younger children are relatively less concerned with others' perceptions of them than older children and adults (e.g., Dunham et al., 2011). Therefore, younger children might generally prioritize epistemic goals, while older children might increasingly weigh social goals. As a consequence, younger children might also be particularly likely to experience the benefits of disagreement on metacognition and learning. However, it is also plausible that even for young children, the metacognitive benefits they obtain from experiencing disagreement vary based on the specific circumstances under which the disagreement is taking place. Hence, in the following sections, I will explore four characteristics of disagreements that might modulate the effects of disagreement on metacognition from early in life.

Cooperative versus competitive disagreement contexts

The disagreements in the studies presented across this dissertation all occurred within contexts of collaborative decision-making. That is, children and the disagreeing other – whether this was an adult confederate or a same-aged peer – were jointly pursuing a shared epistemic goal (such as determining which toys activated a machine or which of two boxes contained the reward). Such collaborative reasoning settings have previously been identified as pivotal for children's cognitive development (see Domberg et al., 2018; Köymen & Tomasello, 2020; Tomasello, 2021) – and the current work has expanded this understanding by suggesting that

encountering disagreements *within* collaborative contexts might be particularly beneficial for children's development, because it motivates them to introspect on their own knowledge.

In contrast, encountering a disagreement under more competitive conditions might be less likely to lead to metacognitive reflection because children may be more driven to by the motivation to "win" the argument than by the epistemic goal to determine the truth and consider the possibility of being wrong (see Dweck, 1986; Dweck & Bempechat, 2017; Fisher & Keil, 2016). Such a motivation to win the argument and outperform the disagreeing, in turn, other may be driven by the social goal to establish (or maintain) a positive personal reputation as someone who is clever or smart or does not make any mistakes. Prior research has shown that children engage in strategic reputation-management from early in life, although it has mostly focused on children's motivation to present themselves as prosocial (see Engelmann & Rapp, 2018; Heymann et al., 2021). Future research should determine if such reputational concerns also extend to the epistemic domain, and if they affect children's willingness to question their own knowledge when presented with disagreement.

Significance of the belief one is disagreeing about

In the current studies, children experienced disagreements about beliefs they had only just acquired, and that thus likely were not particularly important to them. When disagreeing about beliefs that are of greater personal importance to them, children might be less willing to question whether these beliefs may be wrong, because this might conflict with their social goal of maintaining a coherent sense of self (Kahan, 2017). This is suggested by research with adults who often respond defensively or even aggressively when confronted with disagreement about beliefs that constitute a core part of their identity (e.g., Kahan, 2012; Lewandowsky & Oberauer, 2016; Nyhan & Reifler, 2010; Ranalli, 2020; Ranalli & Lagewaard, 2022).

Although children's identities are generally not yet as fully formed as those of adults and their beliefs are not yet as deeply ingrained, it has been argued that the need for self-coherence and identity protection drives children's behavior from infancy (Dweck, 2017). While assessing the significance of various beliefs in children would undoubtedly be challenging, it would be very interesting to determine the relative importance children ascribe to various longer-held beliefs. Then, one could introduce children to disagreement about these beliefs and test if children are more inclined to reflect on beliefs that are less central to their identity compared to those that are more important to them.

Group membership of the disagreeing individuals

In addition to the context of the disagreement and the belief that they are disagreeing about, children's willingness to reflect on their beliefs could also be influenced by who they are disagreeing with. Specifically, research with adults suggests that they are often more willing to consider the possibility that they could be wrong if the person they are disagreeing with is perceived as part of their ingroup, while disagreements with outgroup members trigger more defensive and competitive responses (see, e.g., Kahan et al., 2011). This bias is likely grounded in the social goal to affiliate with one's ingroup and maintain group cohesion (e.g., Tomasello, 2014).

While it is widely acknowledged that children's intergroup attitudes affect their behavior from early in life (e.g., Dunham et al., 2011; Levy & Killen, 2008), we know surprisingly little about whether they also affect children's reasoning; let alone their metacognitive reasoning. Thus, an important step for future research will be to determine whether already young children

might be less willing to metacognitively reflect on their beliefs when disagreeing with an outgroup, compared to an ingroup member. Beyond that, it would be interesting to determine whether children's responses to disagreement with outgroup-members might be further modulated by whether the disagreement is framed as a competition between the two social groups, or whether the disagreeing individuals' mutual goal of determining the truth is highlighted (see Sherif, 2015).

Power dynamics between the disagreeing individuals

Lastly, another factor that could influence the effects of disagreement on children's metacognition are the power dynamics between the disagreeing parties, since already infants are sensitive to power and dominance hierarchies (e.g., Castelain et al., 2016; Fusaro, Corriveau & Harris, 2011; Mascaro & Csibra, 2012). On one hand, when two children of significantly different social statuses disagree, the higher-status child might be more reluctant to engage in metacognitive reflection when confronted with the alternative perspective by the lower-status child. This is because accepting the lower-status child's perspective as equal might conflict with the higher-status child's social goal to maintain their privileged social position (see Fricker, 2007). Schleihauf, Langenhoff, et al., in prep).

However, a significant power differential between the disagreeing parties may also modulate the potential metacognitive benefits for the lower-status child. This is because the lower-status child might simply defer to the belief of the higher-status child without critically engaging with it (see Kruger & Tomasello, 1986; Piaget, 1952). Thus, disagreements involving strong power-differentials might limit the benefits of disagreement for both disagreeing parties, while disagreements in which both partners have equal status and mutually respect each other might incur greater metacognitive – as well as social – benefits.

In summary, my goal in the current section was to begin to outline factors that could plausibly modulate the effects of disagreement even on young children's metacognition. Importantly, the extent to which these factors influence the relation between disagreement and metacognition likely depends on a disagreeing child's age, their cultural and personal background, as well as more or less subtle contextual cues. As such, future research needs to carefully manipulate these factors in children of different ages and cultural backgrounds, and across different contexts to determine which types of disagreement best promote metacognitive reflection and learning.

5.4 Conclusion

Young children are often described as little "know-it-alls" (Hagà & Olson, 2017), who are largely unaware of their own cognitive limitations (e.g., Roebers, 2017). However, by middle childhood, children's metacognitive competence has significantly improved. What drives this striking developmental change? The current dissertation explored a previously overlooked factor underlying this improvement: the social experience of disagreement. In doing so, it brings together two influential theoretical perspectives on cognitive development – constructivist learning theories (e.g., Piaget, 1952), and cultural evolutionary accounts of metacognition (e.g., Heyes et al., 2020) – refines them, and makes them empirically testable.

Across five studies, the current research demonstrated that disagreement prompted young children's metacognition in three key ways: by prompting their reason-giving during joint

decision-making, reducing their overconfidence, and encouraging their rational belief-revision and suspension of judgment. By showing that even 4-year-olds display these sophisticated metacognitive abilities within contexts of social disagreement, this research constitutes a significant departure from earlier work on metacognitive development, which has often been skeptical regarding metacognitive competencies of children that age. Although many questions for future research remain, the current work makes a significant theoretical contribution to understanding the role of social interaction in cognitive development, and suggests exciting possibilities for interventions aimed at fostering learning and nurturing essential citizenship skills.

References

Amemiya, J., Heyman, G. D., & Gerstenberg, T. (2024). Children use disagreement to infer what happened. *Cognition, 250,* 105836.

Amemiya, J., Walker, C. M., & Heyman, G. D. (2021). Children's developing ability to resolve disagreements by integrating perspectives. *Child Development, 92*(6), e1228-e1241.

Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive change by social conflict. *Developmental Psychology, 18*(6), 894–897.

Arendt, H. (1958). *The human condition.* University of Chicago Press.

Asaba, M., & Gweon, H. (2018). Look, I can do it! Young children forego opportunities to teach others to demonstrate their own competence. In *Proceedings of the 40th meeting of the Cognitive Science Society*. London: Cognitive Science Society.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70,* 1–70.

Baer, C., & Kidd, C. (2022). Learning with certainty in childhood. *Trends in Cognitive Sciences, 26*(10), 887-896.

Baer, C., Langenhoff, A.F., Keşşafoğlu, D., Mohtezebsade, W., Kidd, C., Küntay, A., Köymen, B., Engelmann, J.M. (in prep). Does Disagreement Affect Source Memory in English-Speaking and Turkish-Speaking Preschoolers?

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science, 329*(5995), 1081-1085.

Banerjee, P. N., & Tamis-LeMonda, C. S. (2007). Infants' persistence and mothers' teaching as predictors of toddlers' cognitive development. *Infant Behavior and Development, 30,* 479–491.

Barker, C. (2013). Negotiating taste. *Inquiry, 56*(2-3), 240-257.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., & Scheipl, F. (2012). *Package 'lme4'. CRAN. R Foundation for Statistical Computing*, Vienna, Austria.

Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology Monograph, 4*, 1-103.

Bearison, D. J., Magzamen, S., & Filardo, E. K. (1986). Socio-cognitive conflict and cognitive growth in young children. *Merrill-Palmer Quarterly,* 51-72.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*(6), 1641-1660.

Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three-and four-year-olds' spontaneously use others' past performance to guide their learning. *Cognition, 107,* 1018–1034.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, 120*(3), 322-330.

Bornstein, M. H. (2013). *Cultural approaches to parenting.* Psychology Press.

Bridgers, S., Buchsbaum, D., Seiver, E., Griffiths, T. L., & Gopnik, A. (2016). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology, 52,* 9–18.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry, 18*(1), 1-21.

Buehler, F. J., Ghetti, S., & Roebers, C. M. (2023, June 27). *Training Primary School Children's Uncertainty Monitoring.* https://osf.io/preprints/yw476
Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology, 24,* 654–663.
Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition, 3,* 207-220.
Castelain, T., Bernard, S., Van der Henst, J. B., & Mercier, H. (2016). The influence of power and reason on young Maya children's endorsement of testimony. *Developmental Science, 19*(6), 957-966.
Cheek, N. N., Blackman, S. F., & Pronin, E. (2021). Seeing the subjective as objective: People perceive the taste of those they disagree with as biased and wrong. *Journal of Behavioral Decision Making, 34*(2), 167-182.
Chen, E. E., Corriveau, K. H., & Harris, P. L. (2013). Children trust a consensus composed of outgroup members—but do not retain that trust. *Child Development, 84,* 269–282.
Christensen, D. (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass, 4*(5), 756-767.
Coleman, P. T., Deutsch, M., & Marcus, E. C. (Eds.). (2014). *The handbook of conflict resolution: Theory and practice.* John Wiley & Sons.
Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer non-dissenters as informants. *Psychological Science, 20,* 372–377.
Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science, 18,* 957–971.
Crawford, L. (2022). Suspending judgment is something you do. *Episteme, 19*(4), 561-577.
Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A 3-year intervention study. *Journal of Cognition and Development, 15*(2), 363-381.
Danniels, E., & Perlman, M. (2021). Because 'everybody believes in different things': Examining tolerance of divergent preferences, beliefs, and morals in kindergarten students. *Social Development, 30*(4), 973-993.
Danovitch, J. H., Fisher, M., Schroder, H., Hambrick, D. Z., & Moser, J. (2019). Intelligence and neurophysiological markers of error monitoring relate to children's intellectual humility. *Child Development, 90*(3), 924-939.
De Dreu, C. K., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis. *Journal of Applied Psychology, 88*(4), 741.
Dessler, A. E., & Parson, E. A. (2019). *The science and politics of global climate change: A guide to the debate.* Cambridge University Press.
Destan, N., Hembacher, E., Ghetti, S., & Roebers, C. M. (2014). Early metacognitive abilities: The interplay of monitoring and control processes in 5- to 7-year-old children. *Journal of Experimental Child Psychology, 126,* 213–228.
Dewey, J. (1922). *Human Nature and Conduct: An Introduction to Social Psychology.* New York: Modern Library.
Diesendruck, G., & Bloom, P. (2003). How specific is the shape bias? *Child Development, 74,* 168-178.
Dobson, A. J., & Barnett, A. G. (2018). *An introduction to generalized linear models.* CRC press.
Doise, W., & Mugny, G. (1979). Individual and collective conflicts of centrations in cognitive

development. *European Journal of Social Psychology, 9*(1), 105–108.
Mugny, G., & Doise, W. (1978). Socio-cognitive conflict and structure of individual and collective performances. *European Journal of Social Psychology, 8*(2), 181-192.
Doise, W., Mugny, G., & Perret-Clermont, A. N. (1976). Social interaction and cognitive development: Further evidence. *European Journal of Social Psychology.*
Doise, W., Mugny, G., & Perret-Clermont, A. N. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5*(3), 367-383.
Domberg, A., Köymen, B., & Tomasello, M. (2018). Children's reasoning with peers in cooperative and competitive contexts. *British Journal of Developmental Psychology, 36*(1), 64-77.
Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287-312.
Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22,* 271–280.
Dunn, J., & Munn, P. (1987). Development of justification in disputes with mother and sibling. *Developmental Psychology, 23,* 791.
Dunn, J., & Herrera, C. (1997). Conflict resolution with friends, siblings, and mothers: A developmental perspective. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression, 23*(5), 343-357.
Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. *In Advances in Experimental Social Psychology, 44,* Academic Press.
Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development, 82*(3), 793-811.
Dutilh Novaes, C. (2018). The enduring enigma of reason. *Mind & Language, 33*(5), 513-524.
Dweck, C. S. (1986). Motivational processes affecting learning. *American psychologist, 41*(10), 1040–1048.
Dweck, C. S. (2017). From needs to goals and representations: Foundations for a unified theory of motivation, personality, and development. *Psychological review, 124*(6), 689.
Dweck, C. S., & Bempechat, J. (2017). Children's theories of intelligence: Consequences for learning. In *Learning and motivation in the classroom* (pp. 239-256). Routledge.
Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256–273.
Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process?. *Educational Research Review, 1*(1), 3-14.
Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*(1), 5–12.
Engelmann, J. M., & Rapp, D. J. (2018). The influence of reputational concerns on children's prosociality. *Current Opinion in Psychology, 20,* 92-95.
Fawcett, L. M., & Garton, A. F. (2005). The effect of peer collaboration on children's problem-solving ability. *British Journal of Educational Psychology, 75*(2), 157-169.
Felton, M., Garcia-Mila, M., & Gilabert, S. (2009). Deliberation versus dispute: The impact of argumentative discourse goals on learning and reasoning in the science classroom. *Informal Logic, 29,* 417–446.
Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of

learning. *Learning and Instruction, 32,* 1-9.
Fisher, M., & Keil, F. C. (2016). The trajectory of argumentation and its multifaceted functions. *The psychology of argument: cognitive approaches to argumentation and persuasion,* 347-362.
Fisher, M., Knobe, J., Strickland, B., & Keil, F. C. (2017). The Influence of Social Interaction on Intuitions of Objectivity and Subjectivity. *Cognitive Science, 41,* 1119–1134.
Foushee, R. & Srinivasan, M. (2017). Could both be right? Children's and adults' sensitivity to subjectivity in language. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society,* 379–3384. London: Cognitive Science Society.
Flavell J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development,* 24, 15–23.
Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906.
Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology, 1*(4), 324-340.
Foley, M. A., Johnson, M. K., & Raye, C. L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Development, 5,* 1-60.
Foushee, R. & Srinivasan, M. (2017). Could both be right? Children's and adult's sensitivity to subjectivity in language. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society.* London: Cognitive Science Society.
Frances, B. (2014). *Disagreement.* John Wiley & Sons.
Freeman, E. E., Karayanidis, F., & Chalmers, K. A. (2017). Metacognitive monitoring of working memory performance and its relationship to academic achievement in Grade 4 children. *Learning and Individual Differences, 57,* 58-64.
Fricker, M. (2007). Epistemic injustice: *Power and the ethics of knowing.* Oxford University Press.
Friedman, J. (2013). Suspended judgment. *Philosophical Studies, 162,* 165–181.
Friedman, J. (2017). Why suspend judging? *Noûs, 51,* 302–326.
Fusaro, M., Corriveau, K.H., & Harris, P.L. (2011). The good, the strong, and the accurate: preschoolers' evaluations of informant attributes. *Journal of Experimental Child Psychology, 110,* 561–574.
Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science, 23*(8), 931-939.
Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development,* 1366-1371.
Goupil, L., & Kouider, S. (2016). Behavioral and neural indices of metacognitive sensitivity in preverbal infants. *Current Biology, 26*(22), 3038-3045.
Goupil, L., & Kouider, S. (2019). Developing a reflective mind: From core metacognition to explicit self-reflection. Current Directions in *Psychological Science, 28*(4), 403-408.
Goupil, L., & Proust, J. (2023). Curiosity as a metacognitive feeling. *Cognition, 231,* 105325.
Greenfield, P. M., Quiroz, B., & Raeff, C. (2000). Cross-cultural conflict and harmony in the social construction of the child. *New directions for child and adolescent development, 2000,* 93–108.
Greven, C. U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T., & Plomin, R. (2009). More than

just IQ: School achievement is predicted by self-perceived abilities—But for genetic rather than environmental reasons. *Psychological Science*, *20*(6), 753-762.
Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition, 132*(3), 335-341.
Habermas, J. (1962). *The structural transformation of the public sphere.* MIT Press.
Habermas, J. (1997). *Contributions to a discourse theory of law and democracy.* Polity Press.
Hagá, S., & Olson, K. R. (2017a). If I only had a little humility, I would be perfect: Children's and adults' perceptions of intellectually arrogant, humble, and diffident people. *The Journal of Positive Psychology, 12,* 87–98.
Hagá, S., & Olson, K. R. (2017b). Knowing-it-all but still learning: Perceptions of one's own knowledge and belief revision. *Developmental Psychology, 53,* 2319.
Haney, R. E. (1964). The development of scientific attitudes. *The Science Teacher, 31,* 33–35.
Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology, 69,* 251-273.
Haun, D. B., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development, 82,* 1759–1767.
Heiphetz, L., & Young, L. L. (2017). Can only one person be right? the development of objectivism and social preferences regarding widely shared and controversial moral beliefs. *Cognition, 167,* 78–90.
Heiphetz, L., Spelke, E. S., Harris, P. L., & Banaji, M. R. (2013). The development of reasoning about beliefs: Fact, preference, and ideology. *Journal of Experimental Social Psychology, 49*(3), 559-565.
Hembacher, E., & Ghetti, S. (2013). How to bet on a memory: Developmental linkages between subjective recollection and decision making. *Journal of Experimental Child Psychology, 115,* 436–452.
Hembacher, E., deMayo, B., & Frank, M. C. (2020). Children's social information seeking is sensitive to referential ambiguity. *Child Development, 91,* e1178-e1193.
Heyes, C. (2012). Grist and mills: On the cultural origins of cultural learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367,* 2181–2191.
Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture, 2*(1), 8.
Hrdy, S. B. (2009). Allomothers across species, across cultures, and through time. *Substitute parents: Biological and social perspectives on alloparenting in human societies, 3.*
Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony
about the physical world. *Cognitive Psychology, 61,* 248–272.
Jaswal, V. K., Pérez-Edgar, K., Kondrad, R. L., Palmquist, C. M., Cole, C. A., & Cole, C. E. (2014). Can't stop believing: Inhibitory control and resistance to misleading testimony. *Developmental Science, 17,* 965–976.
Kahan, D. M. (2012). Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgment and Decision making, 8,* 407-424.
Kelly, T. (2003). Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research, 66,* 612–640.
Koenig, M. A. (2012). Beyond semantic accuracy: Preschoolers evaluate a speaker's reasons. *Child Development, 83,* 1051–1063.

Hanish, L. D., Martin, C. L., Miller, C. F., Fabes, R. A., DeLay, D., & Updegraff, K. A. (2016). Social harmony in schools: A framework for understanding peer experiences and their effects. In *Handbook of social influences in school contexts* (pp. 48-62). Routledge.

Heyes, C. (2018). *Cognitive gadgets.* Harvard University Press.

Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020). Knowing ourselves together: The cultural origins of metacognition. *Trends in Cognitive Sciences, 24,* 349-362.

Heyman, G. D., Compton, A. M., Amemiya, J., Ahn, S., & Shao, S. (2021). Children's reputation management: Learning to identify what is socially valued and acting upon it. *Current Directions in Psychological Science, 30*(4), 315-320.

Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G., Mitchell, D., & Zeileis, M. A. (2015). *Package 'lmtest'. Testing linear regression models.* https://cran. r-project. org/web/packages/lmtest/lmtest. pdf.

Iwasaki, S., & Kishimoto, R. (2021). Studies of prospective information-seeking in capuchin monkeys, pigeons, and human children. *Comparative Cognition: Commonalities and Diversity,* 255-267.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science, 22,* 129-146.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365-379.

Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research, 14*(2), 147-174.

Kahan, D. M. (2017). Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition. *SSRN Electronic Journal.*

Karataş, K., & Arpacı, İ. (2021). The role of self-directed learning, metacognition, and 21st centuryskills predicting the readiness for online learning. *Contemporary Educational Technology, 13*(3)

Kelly, T. (2005). The epistemic significance of disagreement. *Oxford studies in epistemology, 1*(167-196).

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1322-1337.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one, 7*(5), e36399.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*(2), 211.

Koenig, M. A., & Jaswal, V. K. (2011). Characterizing Children's Expectations About Expertise and Incompetence: Halo or Pitchfork Effects? *Child Development, 82,* 1634–1647.

Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving schools, 19*(3), 267-277.

Köymen, B., & Engelmann, J. M. (2022). Young children rely on gossip when jointly reasoning about whom to believe. *Developmental Psychology, 58,* 1091.

Köymen, B., & Tomasello, M. (2020). The early ontogeny of reason giving. *Child Development Perspectives, 14*, 215–220.
Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121.
Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5),
Kuhn, D. (2014). Theory of mind, metacognition, and reasoning: A life-span perspective. In *Children's reasoning and the mind* (pp. 301-326). Psychology Press.
Kuhn, D. (2015). Thinking together and alone. *Educational researcher, 44*(1), 46-53.
Kuhn, D. (2019). Critical thinking as discourse. *Human Development, 62*(3), 146-164.
Kuhn, D. (2022). Metacognition matters in many ways. *Educational Psychologist, 57*(2), 73-86.
Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science, 22*(4), 545-552.
Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15*(3), 309-328.
Kuhn, T. S. (1962). *The structure of scientific revolutions.* The University of Chicago Press.
Lackey, J. (2010). *Learning from words: Testimony as a source of knowledge.* OUP Oxford.
Lamers, W. S., Boyack, K., Larivière, V., Sugimoto, C. R., van Eck, N. J., Waltman, L., & Murray, D. (2021). Meta-Research: Investigating disagreement in the scientific literature. *Elife, 10*, e72737.
Langenhoff, A. F., Engelmann, J. M., & Srinivasan, M. (2023). Children's developing ability to adjust their beliefs reasonably in light of disagreement. *Child Development, 94*(1), 44-59.
Langenhoff, A. F.*, Regan, S.,* Jacobs, C., Engelmann, J. M., & Srinivasan, M. Children's developing sense of epistemic injustice – evidence from Gujarat, India. *equal contribution as first author
Langenhoff, A. F., Srinvasan, M., & Engelmann, J.M. (2024). Disagreement Reduces Overconfidence and Prompts Curiosity in Young Children. *Child Development*
Langenhoff, A.F.*, Schleihauf, H.*, Herrmann, E., Zeidler, H., & Engelmann, J.M. (in prep). Kikuyu and US children show different developmental patterns of overconfidence. *equal contribution as first authors
Langenhoff, A.F.,Thompson, B.D.,Srinivasan, M. & Engelmann, J.M. (in prep). Disagreement Drives Metacognitive Development.
Lapidow, E., Killeen, I., & Walker, C. M. (2022). Learning to recognize uncertainty vs. recognizing uncertainty to learn: Confidence judgments and exploration decisions in preschoolers. *Developmental Science, 25*(2), e13178.
Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences, 24*(1), 65-78.
Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, K. T., & Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality and Social Psychology Bulletin, 43*(6), 793-813.
Levy, S. R., & Killen, M. (Eds.). (2008). *Intergroup attitudes and relations in childhood through adulthood.* Oxford University Press.
Liquin, E. G., & Gopnik, A. (2022). Children are more exploratory and learn more than adults in an approach-avoid task. *Cognition*, 218, 104940.
Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief

understanding: A training study. *Child Development, 74*(4), 1130-1144.
Deffler, S. A., & Hoyle, R. H. (2017). Cognitive and Interpersonal Features of Intellectual Humility. *Personality and Social Psychology Bulletin, 43,* 793–813.
Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *Emmeans: Estimated marginal means, aka least-squares means.* R package version, 1, 3.
Leonard, J. A., Duckworth, A. L., Schulz, L. E., & Mackey, A. P. (2021). Leveraging cognitive science to foster children's persistence. *Trends in Cognitive Sciences, 25,* 642-644.
Lewandowsky, S., & Oberauer, K. (2016). Motivated rejection of science. *Current Directions in Psychological Science, 25,* 217–222.
Li, L., & Tomasello, M. (2022). Disagreement, justification, and equitable moral judgments: A brief training study. *Journal of Experimental Child Psychology, 223,* 105494.
Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology, 103,* 152–166.
Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate
judgments of learning. *Developmental Psychology, 49,* 1505–1516.
Loftus, E. F., & Pickrell, J.E. (1995). The formation of false memories. *Psychiatric Annals, 25*(12), 720-725.
Lynch, M. P., Johnson, C. R., Sheff, N., & Gunn, H. (2016). *Intellectual humility in public discourse.* IHPD Literature Review. https://humilityandconviction. uconn. edu/wp-content/uploads/sites/1877/2016/09/IHPD-Literature-Review-revised. pdf.
Ma, L., & Ganea, P. A. (2010). Dealing with conflicting information: Young children's reliance on what they see versus what they are told. *Developmental Science, 13,* 151–160.
Ma, Y., Bang, D., Wang, C., Allen, M., Frith, C., Roepstorff, A., & Han, S. (2014). Sociocultural patterning of neural activity during self-reflection. *Social Cognitive and Affective Neuroscience, 9*(1), 73-80.
Mahr, J. B., & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences, 41,* e1.
Mahr, J. B., Mascaro, O., Mercier, H., & Csibra, G. (2021). The effect of disagreement on children's source memory performance. *PLOS ONE, 16,* e0249958.
Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. *Child Development,* 986-992.
Martin, A., Ryan, R. M., & Brooks-Gunn, J. (2013). Longitudinal associations among interest, persistence, supportive parenting, and achievement in early childhood. *Early Childhood Research Quarterly, 28,* 658-667.
Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences, 109*(18), 6862-6867.
Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition, 112*(3), 367-380.
Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach.* Routledge.
Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(2), 57-74.
Mercier, H., Sudo, M., Castelain, T., Bernard, S., & Matsui, T. (2018). Japanese preschoolers'

evaluation of circular and non-circular arguments. *European Journal of Developmental Psychology, 15,* 493–505.
Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*(3), 159-163.
Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and learning, 8,* 19-46.
Michalsky, T., Mevarech, Z. R., & Haibi, L. (2009). Elementary school children reading scientific texts: Effects of metacognitive instruction. *The Journal of Educational Research, 102*(5), 363-376.
Mill, J. S. (1859). *On liberty.*
Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning, 6, 303*-314.
Miosga, N., Schultze, T., Schulz-Hardt, S., & Rakoczy, H. (2020). Selective social belief revision in young children. *Journal of Cognition and Development, 21,* 513–533.
Mokrova, I. L., O'Brien, M., Calkins, S. D., Leerkes, E. M., & Marcovitch, S. (2013). The role of persistence at preschool age in academic skills at kindergarten. *European Journal of Psychology of Education, 28,* 1495–1503.
Moore, D. A. (2022). Overprecision is a property of thinking systems. *Psychological Review.*
Morgan, T. J., Laland, K. N., & Harris, P. L. (2015). The development of adaptive conformity in young children: effects of uncertainty and consensus. *Developmental Science, 18*(4), 511-524.
Moshman, D. (2020). *Reasoning, argumentation, and deliberative democracy.* Routledge.
Moshman, D., & Geil, M. (1998). Collaborative Reasoning: Evidence for Collective Rationality. *Thinking & Reasoning, 4*(3), 231-248.
Nagel, J. (2015). The social value of reasoning in epistemic justification. *Episteme, 12*(2), 297-308.
Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology, 33*(3), 345-359.
Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior, 32,* 303-330.
O'Madagain, C., Helming, K. A., Schmidt, M. F., Shupe, E., Call, J., & Tomasello, M. (2022). Great apes and human children rationally monitor their decisions. *Proceedings of the Royal Society B, 289(1971),* 20212686.
O'Madagain, C. (2019). Is reasoning culturally transmitted? *Teorema: Revista internacional de filosofía, 38,* 107–120.
O'Madagain, C., & Tomasello, M. (2021). Joint attention to mental content and the social origin of reasoning. *Synthese, 198,* 4057-4078.
Oktar, K., & Lombrozo, T. (2022). Mechanisms of Belief Persistence in the Face of Societal Disagreement. *Proceedings of the Annual Meeting of the Cognitive Science Society (44).* London: Cognitive Science Society.
Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328(5977),* 463-466.
Perner, J. (2012). MiniMeta: in search of minimal criteria for metacognition. *Foundations of Metacognition,* 94-116.
Perret-Clermont, A. N. (1980). *Social interaction and cognitive development in children.*

Academic Press.
Piaget, J. (1932). *The Moral Judgment of the Child.* The Free Press, New York.
Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.
Porter, T., & Schumann, K. (2018). Intellectual humility and openness to the opposing view. *Self and Identity, 17*(2), 139–162.
Price, N. (1996). The changing value of children among the Kikuyu of Central Province, Kenya. *Africa, 66*(3), 411-436.
Price, G. F., Ogren, M., & Sandhofer, C. M. (2022). Sorting out emotions: How labels influence emotion categorization. *Developmental Psychology, 58*(9), 1665.
R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.
Ranalli, C. (2020). Deep disagreement and hinge epistemology. *Synthese, 197*(11), 4975-5007.
Ranalli, C., & Lagewaard, T. (2022). Deep Disagreement (Part 1): Theories of Deep Disagreement. *Philosophy compass, 17*(12), e12886.
Rapanta, C., & Felton, M. K. (2021). Learning to Argue Through Dialogue: a Review of Instructional Approaches. *Educational Psychology Review, 34*(2), 477–509.
Rinne, L. F., & Mazzocco, M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PloS one, 9*(7), e98663.
Roderer, T., & Roebers, C. M. (2014). Can you see me thinking (about my answers)? Using eye-tracking to illuminate developmental differences in monitoring and control skills and their relation to performance. *Metacognition and learning, 9*, 1-23.
Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental review, 45*, 31-51.
Roy, W. G. (1984). Class conflict and social change in historical perspective. *Annual Review of Sociology, 10*(1), 483-506.
Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521-562.
Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education, 93*(3), 448-484.
Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology, 20*, 416–420.
Schleihauf, H., Herrmann, E., Fischer, J., & Engelmann, J. M. (2022). How children revise their beliefs in light of reasons. *Child Development, 93*(4), 1072-1089.
Schleihauf, H., Langenhoff, A.F., Herrmann, E., & Engelmann, J.M, in prep. The influence of power on children's reason giving in two diverse societies.
Schneider, W., & Löffler, E. (2016). *The development of metacognitive knowledge in children and adolescents.* In J. Dunlosky & S. K. Tauber (Eds.), Oxford library of psychology. The Oxford handbook of metamemory.
Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right... if they argue together! *Cognition and Instruction, 18*(4), 461-494.
Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development, 23*, 278–290.
Sen, A. (2003). Democracy and its global roots: *Why democratization is not the same as*

westernization. Moral Issues in Global Perspective: Volume I: Moral and Political Theory, 1 208.
Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences, 18*(4), 186-193.
Sherif, M. (2015). *Group conflict and co-operation: Their social psychology.* Psychology Press.
Shneidman, L., Gweon, H., Schulz, L. E., & Woodward, A. L. (2016). Learning from others and spontaneous exploration: A cross-cultural investigation. *Child Development, 87*(3), 723-735.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86,* 420.
Sills, J., Rowse, G., & Emerson, L. M. (2016). The role of collaboration in the cognitive development of young children: a systematic review. *Child: care, health and development, 42*(3), 313-324.
Silver, L., Fetterolf, J., & Connaughton, A. (2021). *Diversity and division in advanced economies.* Pew Research Center. https://www.pewresearch.org/global/2021/10/13/diversity-and-division-in-advanced-economies/
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3.
Sodian B., Thoermer C., Kristen S., Perst H. (2012). Metacognition in infants and young children. In Beran M. J., Brandl J. L., Perner J., Proust J. (Eds.), *Foundations of metacognition* (pp. 119–133). Oxford, England: Oxford University Press.
Stahl, A. E., & Woods, L. (2022). Infants preferentially learn from surprising teachers. *Infancy, 27,* 887-899.
Stanovich, K. (2011). *Rationality and the reflective mind.* Oxford University Press.
Steup, M. & Neta, R. *Epistemology.* The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/spr2024/entries/epistemology
Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology, 44*(3), 187-194.
Tomasello, M. (2020). The ontogenetic foundations of epistemic norms. *Episteme, 17,* 301 315.
Tomasello, M. (2024). An agency-based model of executive and metacognitive regulation. *Frontiers in Developmental Psychology, 2,* 1367381.
Turri, J. (2012). A puzzle about withholding. *The Philosophical Quarterly, 62,* 355–364.
van Loon, M. H., & Roebers, C. M. (2020). Using feedback to improve monitoring judgment accuracy in kindergarten children. *Early Childhood Research Quarterly, 53,* 301-313.
van Loon, M., de Bruin, A., Leppink, J., & Roebers, C. (2017). Why are children overconfident? Developmental differences in the implementation of accessibility cues when judging concept learning. *Journal of Experimental Child Psychology, 158,* 77–94.
Vygotsky, L. (1987). *The collected works of L.S. Vygotsky: Problems of general psychology* (R. Rieber & A. Carlton, Eds.). New York, NY: Plenum. (Original work published 1937)
Wang, C., Oyserman, D., Liu, Q., Li, H., & Han, S. (2013). Accessible cultural mind-set modulates default mode activity: evidence for the culturally situated brain. *Social Neuroscience, 8*(3), 203-216.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23*(1), 63-71.

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., & Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition, 22*(1), 264-271.

Westrum, R. (1989). 14. The psychology of scientific dialogues. *Psychology of Science: Contributions to Metascience, 370.*

Wiebe, S. A., Sheffield, T. D., & Espy, K. A. (2012). Separating the fish from the sharks: A longitudinal study of preschool response inhibition. *Child Development, 83,* 1245-1261.

Wood, D. F. (2003). Problem based learning. *Bmj, 326*(7384), 328-330.

Wu, Y., & Gweon, H. (2021). Preschool-aged children jointly consider others' emotional expressions and prior knowledge to decide when to explore. *Child Development, 92*(3), 862-870.

Yang, Q. T., Sleight, S., Ronfard, S., & Harris, P. L. (2023). Young children's conceptualization of empirical disagreement. *Cognition, 241,* 105627.

Yoon, E. J., MacDonald, K., Asaba, M., Gweon, H., & Frank, M. C. (2018). Balancing informational and social goals in active learning. In *Proceedings of the 40th meeting of the Cognitive Science Society.* London: Cognitive Science Society.

Young, A. G., Alibali, M. W., & Kalish, C. W. (2012). Disagreement and causal learning: Others' hypotheses affect children's evaluations of evidence. *Developmental Psychology, 48,* 1242 – 1253.

Zeman, D. (2019). Faultless disagreement. In *The routledge handbook of philosophy of relativism* (pp. 486-495). Routledge.

Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2019). Cognitive inflexibility predicts extremist attitudes. *Frontiers in Psychology, 10,* 424519.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. Journal of Research in Science Teaching: *The Official Journal of the National Association for Research in Science Teaching, 39*(1), 35-62.